



计算机科学

COMPUTER SCIENCE

基于双向交叉注意力跨域融合的航拍图像伪装目标识别方法

李昂, 章杰元, 刘逊韵

引用本文

李昂, 章杰元, 刘逊韵. 基于双向交叉注意力跨域融合的航拍图像伪装目标识别方法[J]. 计算机科学, 2026, 53(1): 173-179.

LI Ang, ZHANG Jieyuan, LIU Xunyun. [Camouflaged Object Detection for Aerial Images Based on Bidirectional Cross-attentionCross-domain Fusion](#) [J]. Computer Science, 2026, 53(1): 173-179.

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)

[面向全天候多场景的多模态融合目标检测方法](#)

Multi-modal Fusion Based Object Detection for All-day and Multi-scenario Environments

计算机科学, 2025, 52(11A): 241100137-10. <https://doi.org/10.11896/jsjcx.241100137>

[基于运动模式与时频域融合的行人轨迹预测](#)

Pedestrian Trajectory Prediction Based on Motion Patterns and Time-Frequency Domain Fusion

计算机科学, 2025, 52(7): 92-102. <https://doi.org/10.11896/jsjcx.250200011>

[面向无人机航拍图像的YOLOv8目标检测改进算法](#)

An Improved YOLOv8 Object Detection Algorithm for UAV Aerial Images

计算机科学, 2025, 52(4): 202-211. <https://doi.org/10.11896/jsjcx.240500042>

[多源异构数据渐进式融合的虚假新闻检测](#)

Multi-source Heterogeneous Data Progressive Fusion for Fake News Detection

计算机科学, 2024, 51(11): 30-38. <https://doi.org/10.11896/jsjcx.240700004>

[航拍绝缘子图像缺陷智能识别算法研究](#)

Study on Intelligent Defect Recognition Algorithm of Aerial Insulator Image

计算机科学, 2024, 51(6A): 230700172-5. <https://doi.org/10.11896/jsjcx.230700172>

基于双向交叉注意力跨域融合的航拍图像伪装目标识别方法

李昂 章杰元 刘逊韵

军事科学院 北京 100091

(angli.cs@outlook.com)

摘要 针对航拍图像中伪装目标与环境高度融合以及实时性要求高的挑战,提出了一种基于双向交叉注意力跨域融合的伪装目标识别方法。首先,提出跨域双分支特征提取网络,分别从 RGB 域和频域进行提取特征,从而增强对低对比度图像的特征提取能力。同时,通过设计双向交叉注意力融合模块,在多个尺度上将频域特征与 RGB 特征进行基于注意力机制的双向特征融合,从而有效提升网络表征能力。实验结果表明,相比其他代表性算法,所提出的方法在目标识别准确性与实时性上达到了更优的平衡。

关键词 航拍图像;伪装目标识别;频域特征;双向交叉注意力

中图分类号 TP391.41

Camouflaged Object Detection for Aerial Images Based on Bidirectional Cross-attention Cross-domain Fusion

LI Ang, ZHANG Jieyuan and LIU Xunyun

Academy of Military Sciences, Beijing 100091, China

Abstract To address the challenges of highly integration with the environment and the high demand for real-time performance, this paper proposes a camouflaged object detection model for aerial images using bidirectional cross-attention cross-domain fusion. Firstly, a feature extraction network with two branches is constructed to extract features from both RGB and frequency domain. Simultaneously, frequency features and RGB features are crossly fused at multiple scales using bidirectional cross-attention fusion modules, effectively improving the network's representational capacity. Experimental results show that the proposed model achieves a better balance between target recognition accuracy and real-time performance, compared to other representative models.

Keywords Aerial images, Camouflaged object detection, Frequency domain, Bidirectional cross-attention

1 引言

近年来,随着无人机技术的迅速发展,航拍图像在侦察和监控领域得到了广泛应用。然而,在实际环境中,无源伪装技术不仅能有效隐藏目标,而且成本低廉,应用灵活,给侦察识别带来了极大挑战^[1-2]。伪装目标通常与背景高度融合,特征信息不明显,尤其是在航拍图像中,目标物体尺寸较小,尺度特征多样且细节信息不清晰,这对识别的准确性提出了更高的要求^[3-4]。

当前该领域的研究方法主要分为传统图像处理技术和基于深度学习的伪装目标识别方法。传统方法通常依赖像素级的灰度或颜色差异分析,如 Canny 边缘检测^[5]和阈值分割^[6]等技术,并且传统方法通常基于手动构造特征进行识别,准确性较低。近年来,基于深度学习的方法通过多层非线性变换自动提取图像的高级特征表示,显著提升了伪装目标识别的精度,代表性工作包括 FCTNet^[7], UGTR^[8]和 EGNet^[9]等模型。

然而,航拍图像伪装目标识别任务仍然存在以下问题:1)航拍图像中伪装目标与环境融合度高且图像细节不清晰,对识别的准确性带来更大挑战;2)航拍图像伪装目标识别需要快速获取识别结果,同时航拍无人机通常仅搭载有限的硬件计算资源,这对识别模型的计算实时性提出了更高要求^[10]。

针对上述问题,本文提出了一种双向交叉注意力跨域融合网络(Bidirectional Cross-attention Cross-domain Fusion Network, BCCFN)。首先,将频域特征引入伪装目标识别任务中,提出结合频域和 RGB 域的双分支特征提取网络。通过运用频域特征提取模块,将 RGB 图像映射到频域并作为识别模型的额外输入,为提取隐蔽特征信息提供了支撑。其次,通过利用双向交叉注意力融合模块,在多尺度上融合频域特征与 RGB 特征,促进两种特征的互补与协同。具体而言,双向交叉注意力融合模块分别从 RGB 到频域和从频域到 RGB 两个方向进行特征交互,利用注意力权重对特征进行加权融合,从而增强模型对伪装目标的细节和全局结构的捕捉能力。

到稿日期:2025-03-03 返修日期:2025-06-22

基金项目:国家自然科学基金青年科学基金(62206311)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China(62206311).

通信作者:刘逊韵(xunyunliu@outlook.com)

最后,将迁移学习策略应用于航拍图像伪装目标识别任务中,进一步提高所提方法在野生场景下的泛化能力。本文主要贡献如下:

- 1) 提出跨域双分支特征提取网络,有效结合频域和 RGB 域特征,增强网络对低质量图像的特征提取能力;
- 2) 提出双向交叉注意力融合模块,实现了频域特征与 RGB 特征的多尺度融合,有效提升了模型对复杂伪装目标的识别性能;
- 3) 提出双向交叉注意力跨域融合网络,其在真实数据集上达到了识别准确性与识别实时性的良好平衡。

2 相关工作

早期研究主要基于传统图像识别技术,通过分析像素间的灰度或颜色差异来实现目标识别。例如,Anjar 等^[5]提出的基于 Canny 边缘检测模型的目标识别方法,能够在保持边缘定位精度的同时有效提取目标轮廓;Otsu^[6]采用全局阈值识别技术,在目标与背景分布较为明显时表现出较好的识别效果。然而,传统方法在面对复杂光照条件、多样化背景以及多变的伪装手段时,识别效果往往欠佳,容易出现误识别或漏识别现象^[11]。

随着深度学习技术的快速发展,基于深度学习的伪装目

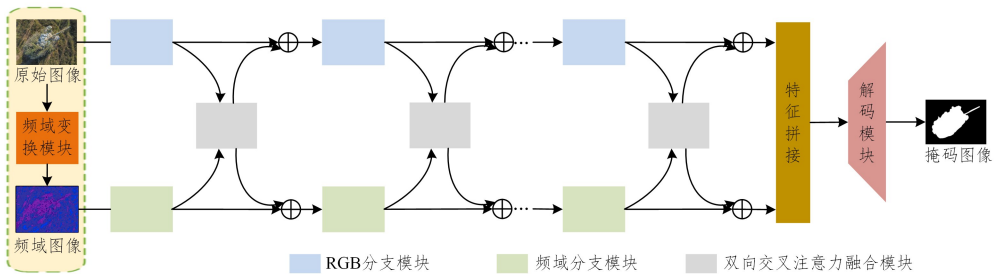


图1 双向交叉注意力跨域融合网络结构

Fig. 1 Structure of bidirectional cross attention cross domain fusion network

网络的特征提取部分由两个平行的分支构成,分别是 RGB 分支和频域分支。在两个分支进行特征提取的过程中,网络在不同深度上引入了多个双向交叉注意力融合模块。该模块通过计算 RGB 特征和频域特征之间的亲和矩阵,促进两种特征在不同域之间的互补与协同。具体而言,双向交叉注意力融合模块分别从 RGB 到频域和从频域到 RGB 两个方向进行特征交互,利用注意力权重对特征进行加权融合,从而增强模型对伪装目标的细节和全局结构的捕捉能力。

最后,网络将 RGB 分支和频域分支提取的多尺度特征进行融合,并通过解码器生成最终的二值分割掩码。解码器采用上采样和跳跃连接结构,逐步恢复特征图的空间分辨率,确保输出的掩码能够精确反映伪装目标的位置和形状。

3.2 跨域双分支特征提取网络

跨域双分支特征提取网络由 RGB 分支与频域分支构成。RGB 分支基于 EfficientNetV2^[14] 构建,用于从 RGB 图像中提取空间特征。EfficientNetV2 通过轻量化卷积模块实现了高效的特征提取,能够在保证模型性能的同时显著降低计算复杂度。具体而言,RGB 分支使用了 EfficientNetV2 的前 4 个阶段(Stage 1 至 Stage 4),其中 Stage 1 至 Stage 3 主要采用

标识方法逐渐成为主流。这类方法通过多层非线性变换自动学习图像中的高级特征表示,能够实现像素级别的分类。PraNet^[12] 通过聚合高层特征生成全局语义图,并利用逆向注意力模块挖掘边界线索,显著提高了目标分割的准确性。Mei 等^[13] 提出的 PFNet,通过模拟捕食过程中的定位和识别阶段,利用定位模块和聚焦模块逐步细化伪装目标的分割结果,有效地提升了伪装目标的分割性能。Yang 等^[8] 进一步提出 UGTR,将贝叶斯学习与 Transformer 推理相结合,通过量化不确定性并利用 Transformer 推理处理不确定区域,进一步提高了伪装目标检测的准确性。此外,Zhao 等^[9] 提出的 EGNNet 通过显式建模边缘信息与目标信息的互补性,利用边缘特征引导目标特征,显著改善了显著目标检测中的边界定位问题。这些研究为伪装目标识别提供了新的视角和方法,特别是在处理复杂背景和多样化的伪装策略方面展示了强大的潜力。

3 双向交叉注意力跨域融合网络

3.1 网络框架

图 1 给出了 BCCFN 的整体框架。该网络以 RGB 图像作为输入,输出为伪装目标的二值分割掩码,旨在实现对航拍图像中伪装目标精确识别的同时满足实时性要求。

Fused-MBConv 模块,该模块通过将 1×1 升维卷积与 3×3 深度卷积融合为单一的 3×3 卷积结构,显著降低了计算开销;Stage 4 采用 MBConv 模块,该模块则通过引入注意力机制和膨胀因子,灵活调节模型容量,以适应不同的资源约束。

频域分支主要针对航拍目标伪装在 RGB 空间中与背景高度融合、边缘模糊且对比度极低的问题^[15]。该分支首先将输入的 RGB 图像通过频域特征提取模块映射到频域,得到频域特征表示,随后采用与 RGB 分支相同的网络结构进行后续特征提取。具体而言,本文提出一种基于离散余弦变换(Discrete Cosine Transform, DCT)的频域特征提取模块。DCT 作为一种高效的频域转换手段,能够在频率维度高效编码图像信号,尤其擅长于凸显图像的边缘与纹理细节。如图 2 所示,输入的 RGB 图像首先被转换到 YCbCr 色彩空间,然后分割成一系列 8×8 的向量,每个向量均进行 DCT 变换,从而将图像信息转换到频率域。DCT 是将图像从空间域转换到频域的关键步骤,其数学表达式为:

$$X_{\text{DCT}}(u, v) = \frac{2}{N} C(u) C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} X(x, y) \quad (1)$$

其中, $X_{\text{DCT}}(u, v)$ 表示变换后频率坐标 (u, v) 处的系数; $X(x,$

y)是原始图像在空间坐标 (x, y) 处的像素值; N 是图像高度; $C(u)$ 和 $C(v)$ 是量化因子。

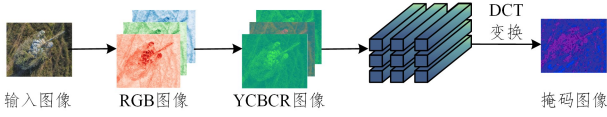


图2 频域变换

Fig. 2 Frequency domain transformation

RGB图像首先转换为YCbCr色彩空间,分离亮度分量(Y)与色度分量(Cb, Cr)。亮度通道Y被分割为非重叠的 8×8 块,每块独立进行离散余弦变换(DCT),生成频域特征块。所有频域特征块重构为与原图尺寸一致的特征图,确保空间对齐。频域特征图经 1×1 卷积调整通道数,与RGB图像通道维度对齐。

通过计算图像中每个像素点与正余弦基函数的内积,得到频域图像。这些系数反映了图像在不同频率上的能量分布,高频系数对应图像中的快速变化部分(如边缘和细节),低频系数对应慢速变化的部分(如大面积的均匀颜色区域)。

3.3 双向交叉注意力融合模块

为了应对航拍图像中伪装目标识别任务中目标与背景高度融合、边缘模糊及对对比度极低所带来的特征提取挑战,提出了一种双向交叉注意力融合模块。图3展示了双向交叉注意力融合模块的具体结构,该模块旨在通过对RGB特征和频域特征进行交叉融合,促进信息在RGB域和频域之间的互补与协同,从而确保模型能够同时捕获目标的细节特征和全局结构,增强模型的表征能力。

通过双向交叉注意力机制实现RGB域与频域特征的多尺度互补融合。其核心设计包含两个交互方向的注意力计算:1)频域到RGB的注意力计算,频域特征 F_i 作为查询(Query),RGB特征 R_i 作为键(Key)和值(Value),生成频域引导的注意力权重,强化RGB特征中与伪装目标相关的全局结构信息;2)RGB到频域的注意力计算,RGB特征 R_i 作为查询(Query),频域特征 F_i 作为键(Key)和值(Value),生成空间引导的注意力权重,增强频域特征中对边缘和纹理的局部细节表征。

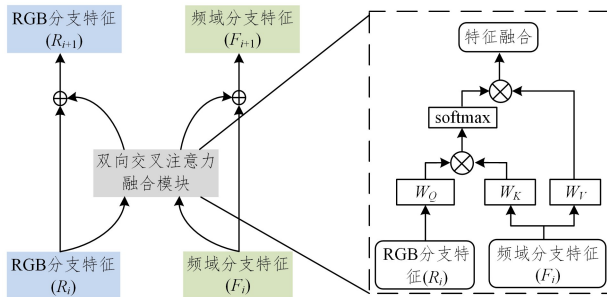


图3 双向交叉注意力跨域融合模块

Fig. 3 Bidirectional cross-attention cross-domain fusion module

这两个过程分别通过计算频域特征和RGB特征之间的亲和矩阵,并利用这些矩阵更新相应的特征表示来实现。为了进一步说明RGB分支特征和频域分支特征的交互过程,

式(2)展示了从频域到RGB域的注意力模块的计算过程。

$$R_{i+1} = R_i + A_{F_i \rightarrow R_i} \cdot F_i W_{rv} \quad (2)$$

其中, W_{rv} 是值嵌入层的学习参数,而 $A_{F_i \rightarrow R_i}$ 是从频域到RGB的亲亲和矩阵,可通过以下方式计算:

$$A_{F_i \rightarrow R_i} = \text{softmax} \left(\frac{R_i W_{rq} \cdot (F_i W_{rk})^T}{\sqrt{d}} \right) \quad (3)$$

其中, W_{rv} , W_{rq} 和 W_{rk} 是线性投影层的学习参数,用于将原始特征映射到查询和键空间。

从RGB域到频域的注意力模块的具体计算式如下:

$$F_{i+1} = F_i + A_{R_i \rightarrow F_i} \cdot F_i W_{fv} \quad (4)$$

$$A_{R_i \rightarrow F_i} = \text{softmax} \left(\frac{F_i W_{fq} \cdot (R_i W_{fk})^T}{\sqrt{d}} \right) \quad (5)$$

其中, W_{fv} , W_{fq} 和 W_{fk} 是3个线性投影层的学习参数, $A_{R_i \rightarrow F_i}$ 是从RGB域到频域的亲亲和矩阵。

通过双向交叉注意力融合模块,频域特征和RGB特征进行了深度交互与融合,有效提升了模型对伪装目标的特征提取能力。该模块关注那些在单个域内可能被忽视的重要特征,进而提高伪装目标识别的准确性和鲁棒性。此外,双向交叉注意力融合模块还能够增强模型处理复杂场景的能力,使得即使是在低对比度和高噪声条件下也能保持较高的识别性能。

为了进一步提升特征提取的效果,网络在不同深度上引入了多个双向交叉注意力融合模块。具体而言,在每个EfficientNetV2阶段(Stage 1至Stage 4)之间均插入了该模块。这种设计使得网络能够在多个尺度上进行特征融合,有效增强了模型对伪装目标的特征提取能力。通过在不同阶段之间进行特征交互,RGB特征和频域特征得以充分融合,从而提升了模型对复杂伪装目标的识别性能。

为了进一步提升模型的推理速度并减少计算资源的消耗,对提出的BCCFN模型进行了剪枝处理。在剪枝过程中,采用通道剪枝策略,基于L1-norm对EfficientNetV2中各MBCConv层的通道进行排序,剪枝率设为30%。

3.4 损失函数

本网络训练使用的损失函数由3部分组成,分别为RGB域的二进制交叉熵损失 L_{bce} 、RGB域的交并比损失 L_{iou} 以及频域损失 L_f 。总的损失函数 L_{total} 可以表示为:

$$L_{total} = \alpha L_{bce} + \beta L_{iou} + \gamma L_f \quad (6)$$

其中, α , β 和 γ 分别是 L_{bce} 、 L_{iou} 和 L_f 的权重系数,用于平衡不同损失函数对模型训练的贡献,权重系数分别取值为 $\alpha = 0.5$, $\beta = 0.3$ 和 $\gamma = 0.2$,以确保模型在训练过程中能够同时关注像素级分类、区域一致性以及频域特征的学习。

二进制交叉熵损失 L_{bce} 具体如下:

$$L_{bce}(P_i, M) = -\frac{1}{N} \sum_{n=1}^N \omega_n [M_n \log P_{in} + (1 - M_n) \log (1 - P_{in})] \quad (7)$$

其中, P_i 是模型在第 i 阶段的预测结果, M 是真实掩模, ω_n 是权重, N 是像素总数, P_{in} 和 M_n 分别代表预测掩码和真实掩码中第 n 个像素的值。

交并比损失 L_{iou} 具体如下:

$$L_{\text{iou}}(P_i, M) = 1 - \frac{\sum_{n=1}^N \omega_n \min(P_{in}, M_n)}{\sum_{n=1}^N \omega_n (\max(P_{in}, M_n) - \min(P_{in}, M_n) + M_n)} \quad (8)$$

频域损失 L_f 通过计算输入 RGB 图像、预测掩码、真实掩码经过 DCT 后的差异平方和的 L2 范数, 来确保模型训练从频域维度提供监督。 L_f 具体计算如下:

$$L_f(Y, M, x) = \frac{\|DCT(x \otimes Y) - DCT(x \otimes M)\|_2}{q} \quad (9)$$

其中, x 为输入的 RGB 图像, Y 是预测的掩码, M 是真实的掩码, \otimes 表示乘积操作, q 表示表示归一化损失。

4 实验与分析

4.1 实验设置

4.1.1 模型训练设置

实验采用 Adam 优化器进行模型训练, 初始学习率设为 1×10^{-4} , 衰减系数为 0.1。训练 200 个轮次 (Epoch), 批量大小 (Batch Size) 为 32, 输入图像尺寸为 416×416 。对 BCCFN 模型进行了剪枝处理, 以降低模型的复杂度和减小计算量。实验采用的开源数据集包括 CAMO 以及 MHCD2022^[16]。其中, CAMO 覆盖了多种自然环境下的迷彩伪装物体图像, 而 MHCD2022 是专门针对军事伪装目标设计的数据集。表 1 列出了所用数据集的训练集和测试集分布情况。

表 1 数据集分布

Table 1 Distribution of training and testing datasets

| Dataset | Training set | Test set | Total |
|----------|--------------|----------|-------|
| CAMO | 2026 | 250 | 2276 |
| MHCD2022 | 1158 | 497 | 1655 |

4.1.2 硬件环境设置

实验平台为 Ubuntu 20.04, 内存为 64 GB; CPU 为 Intel Core i9-9900K 16 核心处理器; GPU 为 NVIDIA GeForce RTX 3090 24GB 显存; 基于 PyTorch 1.10.0, 使用 CUDA 11.3 进行计算加速。

4.1.3 对比模型设置

为了验证所提出方法的有效性, 选择了 5 种代表性的伪装目标识别模型进行对比实验。

1) R-MGL^[17]: 通过相互图学习框架将图像分解为两个特定任务的特征图, 分别用于粗略定位和精确捕捉边界细节, 并利用图推理模块优化特征表示。

2) Rank-Net^[18]: 引入对伪装对象进行排名的任务, 通过定位、分割和排名 3 个子任务的联合学习, 揭示伪装对象的可检测性。

3) JCOD^[19]: 专注于同时检测图像中的显著对象和伪装对象, 通过不确定性感知学习框架处理两项任务的内在矛盾属性, 提升模型对复杂环境的适应性和鲁棒性。

4) SINet^[20]: 通过搜索和识别两个阶段实现高效的伪装对象检测, 结合纹理增强模块、邻域连接解码器和分组反转注意力模块, 实现快速准确的搜索与识别。

5) FSPNet^[21]: 通过结合非局部令牌增强模块 (NL-TEM) 和特征收缩解码器 (FSD), 分别优化局部特征表示与特征聚

合能力, 从而提升伪装物体检测的性能。

为了确保实验的公平性, 上述所有对比模型均采用开源代码, 并使用相同的 CAMO 数据集进行预训练, 然后在 MHCD2022 数据集上进行微调。BCCFN 网络与其他 5 种模型规模对比结果如表 2 所列。

表 2 不同模型规模的对比

Table 2 Comparison of different model sizes

| Model | Model size | Parameter size | FLOPs |
|----------|----------------------|----------------------|-------|
| R-MGL | 3.8806×10^8 | 1.0379×10^8 | 27.22 |
| Rank-Net | 4.5341×10^8 | 1.3464×10^8 | 35.81 |
| JCOD | 7.6292×10^8 | 1.8305×10^8 | 28.17 |
| SINet | 4.6435×10^8 | 1.1342×10^8 | 21.39 |
| FSPNet | 9.9400×10^8 | 2.9617×10^8 | 48.10 |
| BCCFN | 8.7330×10^7 | 3.3080×10^7 | 11.36 |

经过剪枝处理后, BCCFN 的模型大小和参数量分别为 8.7330×10^7 和 3.3080×10^7 , 相较于其他模型, 平均减少了约 82.00% 的模型大小和约 74.03% 的参数量。FLOPs 较现有模型平均至少减少 46.89%, 表明 BCCFN 在资源受限的设备上具有更好的适用性。

4.1.4 评价指标设置

实验采用以下 4 个广泛在伪装目标识别任务中使用的评价指标。

1) Structure-measure^[22] (S_s): 结合区域感知的精确性和结构相关信息, 旨在同时考虑真阳性率和真阴性率, 从而提供对识别质量的整体评价。

2) Mean E-measure^[23] (E_e): 考虑每个像素位置的重要性, 通过精确度和召回率的加权调和平均来衡量预测结果与真实标注之间的符合程度。

3) Mean Absolute Error^[24] (MAE): 直接测量预测值与真实值之间的差距。

4) Dice^[25]: 重合率是衡量识别性能好坏的合理有效的指标, 是两个物体重合度的一种测量, 表示两个物体相交的面积占总面积的比值。

4.2 实验结果分析

4.2.1 双向交叉注意力跨域融合网络验证

为表明 BCCFN 网络的有效性, 将其与其他 4 种代表性模型进行了对比。表 3 列出了不同模型在 MHCD2022 数据集上的性能对比结果。

表 3 MHCD2022 数据集上的定量对比

Table 3 Qualitative comparison on the MHCD2022 dataset

| Model | S_s | E_e | MAE | Dice | Average inference time/ms |
|----------|--------------|--------------|--------------|--------------|---------------------------|
| R-MGL | 0.651 | 0.702 | 0.203 | 0.655 | 163.6 |
| Rank-Net | 0.664 | 0.692 | 0.195 | 0.568 | 188.5 |
| JCOD | 0.607 | 0.685 | 0.214 | 0.681 | 164.0 |
| SINet | 0.679 | 0.741 | 0.179 | 0.694 | 139.6 |
| FSPNet | 0.718 | 0.752 | 0.168 | 0.728 | 112.4 |
| BCCFN | 0.763 | 0.781 | 0.156 | 0.764 | 85.3 |

在 MHCD2022 数据集上的实验结果表明, BCCFN 模型在伪装目标识别任务中表现出显著优势。BCCFN 与现有最优模型 FSPNet 相比, S_s 提升了 6.3%, E_e 提升了 3.9%, MAE 降低了 7.1%, Dice 系数提高了 4.9%。同时, BCCFN 的平均推理时间较 FSPNet 缩短了 24.1%, 表明其在保持高精度的

同时显著优化了计算效率。相较于传统模型(如 SINet 和 JCOD 等),BCCFN 在各项指标上均实现了至少 12.4% 的性能提升,验证了跨域双分支特征提取与双向注意力融合模块的有效性。

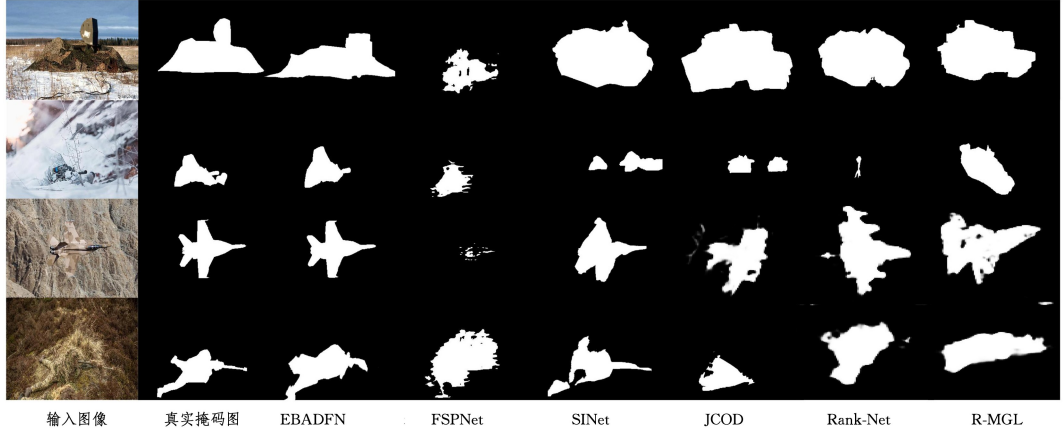


图 4 MHCD2022 数据集上的定性对比

Fig. 4 Qualitative comparison on the MHCD2022 dataset

综上所述,BCCFN 模型通过频域特征提取、双向交叉注意力融合模块,显著提升了伪装目标识别的精度和效率,尤其在硬件资源受限的无人机平台上具有广阔的应用前景。

4.2.2 基于野生测试数据集的验证

为了进一步验证 BCCFN 模型在伪装目标识别上的性能,将从网络上搜集的 500 张包含伪装目标的航拍图片作为野生测试数据集进行实验。

表 4 列出了不同模型在野生测试数据集上的定量对比结果,可以看出 BCCFN 模型表现最为优异。BCCFN 在 S_e 上至少提升了 3.5%,在 E_e 上至少提升了 1.9%,同时 MAE 至少降低了 2.3%,Dice 系数至少提升了 3.1%。实验结果表明,BCCFN 对于目标与背景高度融合的鲁棒性显著优于其他模型,进一步验证了频域特征与注意力机制

图 4 展示了不同模型在 MHCD2022 数据集上的识别结果。BCCFN 在识别伪装目标时表现出更高的准确性和细节捕捉能力,尤其是在目标与背景高度融合的情况下,其能够更准确地定位目标。

在复杂环境中的协同作用。

表 4 野生测试数据集上的定量对比

Table 4 Quantitative comparison on the wild test dataset

| Model | S_e | E_e | MAE | Dice | Average inference time/ms |
|----------|--------------|--------------|--------------|--------------|---------------------------|
| R-MGL | 0.613 | 0.682 | 0.213 | 0.645 | 165.6 |
| Rank-Net | 0.644 | 0.632 | 0.205 | 0.588 | 190.5 |
| JCOD | 0.587 | 0.721 | 0.189 | 0.645 | 145.7 |
| SINet | 0.659 | 0.693 | 0.197 | 0.684 | 161.6 |
| FSPNet | 0.685 | 0.738 | 0.175 | 0.709 | 108.9 |
| BCCFN | 0.709 | 0.752 | 0.171 | 0.731 | 87.3 |

图 5 展示了该野生测试数据集上不同模型的定性对比结果。可以看出,BCCFN 模型在小样本数据集上的识别效果明显优于其他模型。

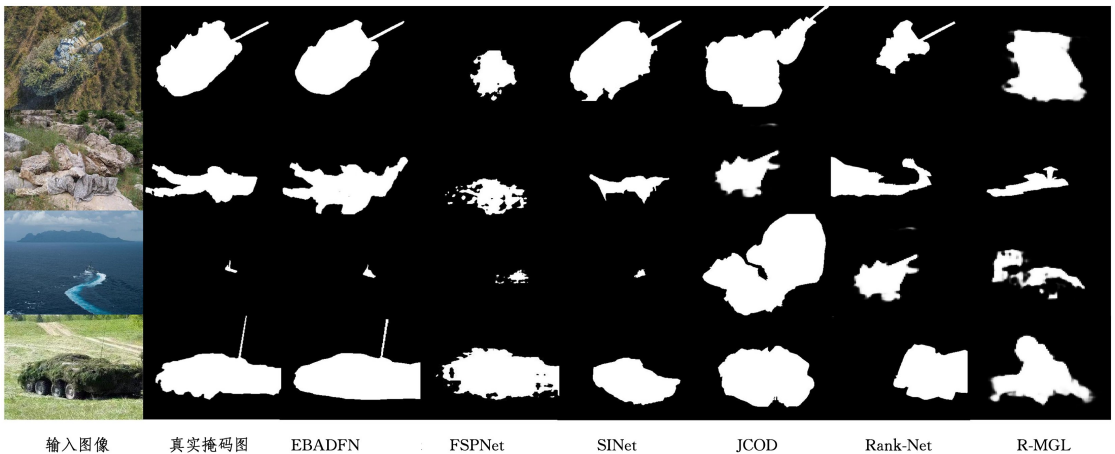


图 5 野生测试数据集上的定性对比

Fig. 5 Qualitative comparison on the wild test dataset

4.2.3 消融实验

本节通过消融实验进一步比较不同 BCCFN 网络变种的性能。所有网络变种均基于 CAMO 数据集进行预训练,并在 MHCD2022 数据集上进行微调。具体实验设置如下:BCCFN

包含 RGB 流、频域流以及双向交叉注意力融合模块;BCCFN-RGB 仅包含 RGB 流,不包含频域流和双向交叉注意力融合模块;BCCFN-RGB-Freq 包含 RGB 流和频域流,但不包含双向交叉注意力融合模块。表 5 列出了 MHCD2022 测试数据

集上的实验对比结果,验证了不同模型变种的性能差异。

表5 消融实验的定量对比

Table 5 Quantitative comparison of ablation experiments

| Model | S_a | E_φ | MAE | Dice | Average inference time/ms |
|----------------|--------------|--------------|--------------|--------------|---------------------------|
| BCCFN-RGB | 0.621 | 0.630 | 0.186 | 0.737 | 88.2 |
| BCCFN-RGB-Freq | 0.691 | 0.754 | 0.171 | 0.728 | 84.4 |
| BCCFN | 0.763 | 0.781 | 0.156 | 0.764 | 85.3 |

从表5可以看出,BCCFN-RGB-Freq在 S_a 和MAE指标上分别比BCCFN-RGB提升了11.3%和8.1%,表明频域特征能够有效增强网络对伪装目标的识别能力。对比BCCFN-RGB-Freq和BCCFN,可验证双向交叉注意力融合模块进一步提升了网络性能。BCCFN的 S_a 、 E_φ 、MAE和Dice均优于BCCFN-RGB-Freq,尤其是 S_a 和MAE分别提升了10.4%和8.8%,表明双向交叉注意力融合模块能够有效融合RGB和频域特征,提升网络表征能力。

图6展示了不同网络变种在MHCD2022数据集上的定性对比结果。BCCFN-RGB虽然能够识别出部分伪装目标,但在目标边缘和细节的捕捉上存在不足,尤其是在目标与背景高度融合的区域,容易出现误识别或漏识别的现象。当引入频域流后,BCCFN-RGB-Freq在目标边缘的识别上有了明显的改善,能够更清晰地勾勒出目标的轮廓,这表明频域特征对低对比度图像的特征提取起到了显著的增强作用,有效弥补了RGB流在细节捕捉上的不足。而完整的BCCFN模型进一步引入双向交叉注意力融合模块后,在目标识别的准确性和完整性上表现最为出色。它不仅能够精准地定位伪装目标的位置,还能完整地分割出目标的形状,即使在复杂背景和与背景高度融合的情况下,也能保持较高的识别精度。这说明双向交叉注意力融合模块能够有效地融合RGB和频域特征,使模型能够更好地捕捉到伪装目标的细节和全局结构信息,从而显著提升了模型的表征能力和识别性能。

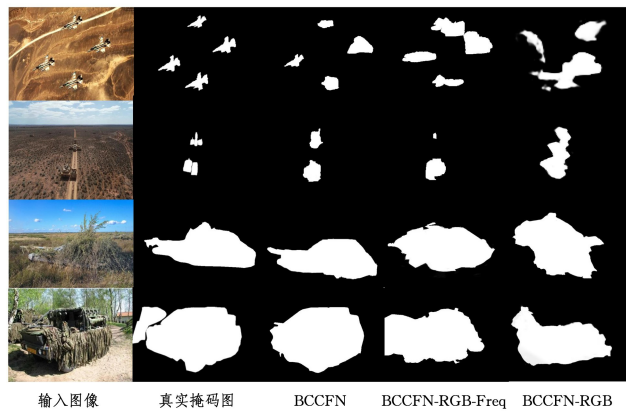


图6 消融实验定性对比

Fig. 6 Qualitative comparison of ablation experiments

结束语 本文提出了一种双向交叉注意力跨域融合网络,结合了频域特征提取方法和双向交叉注意力融合模块,用于提高航拍图像中伪装目标的识别精度。该方法在识别精度、模型参数和推理速度方面均表现出显著优势。引入频域特征提取增强了模型对低对比度图像的特征提取能力,而双

向交叉注意力融合模块融合了RGB和频域特征,提升了网络的表征能力。此外,迁移学习策略有效解决了训练数据不足的问题,提高了模型在小样本情况下的泛化能力。

实验结果表明,所提出的方法在多个公开数据集上取得了优异的识别性能,并在野生数据集上展现了良好的鲁棒性。该方法尤其适用于资源受限的无人机平台,在军事侦察和监控等领域展现出广阔的应用前景。

本文方法尽管在伪装目标识别任务中取得了显著的进展,但仍存在一些问题和挑战。

1)多场景泛化能力:本文方法尽管在多个数据集上表现良好,但在极端环境(如极端光照条件、恶劣天气等)下的泛化能力仍有待提升。未来的研究可以进一步优化模型,尤其是提升其在低光照、雾霾、雨雪等复杂场景下的鲁棒性,使其能够适应更多多样化的环境条件。

2)多模态融合:本文主要基于RGB图像和频域特征进行伪装目标识别,未来可以探索更多模态的数据融合,如红外图像、深度信息等。多模态数据的融合有望进一步提升模型对伪装目标的识别能力。

参考文献

- [1] ZHU L, WANG X, KE Z, et al. BiFormer: vision transformer with bi-level routing attention[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2023: 10323-10333.
- [2] JIANG L, YUAN B, DU J, et al. MFFSODNet: Multi-Scale Feature Fusion Small Object Detection Network for UAV Aerial Images[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-14.
- [3] ZHAO J, ZHANG B, WANG G, et al. Spectral Camouflage Characteristics and Recognition Ability of Targets Based on Visible/Near-Infrared Hyperspectral Images[J]. Photonics, 2022, 9: 957-957.
- [4] DU W C, YU H, ZENG X J, et al. High resolution single-photon imaging for recognition of camouflaged target[C]// Seventh Symposium on Novel Photoelectronic Detection Technology and Applications, 2021.
- [5] ANJAR W, DENI S R, SILFIA A, et al. Combination of Sobel+Prewitt Edge Detection Method with Roberts+Canny on Passion Flower Image Identification [J]. Journal of Physics: Conference Series, 2021, 1933(1): 012037.
- [6] OTSU N. A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [7] KONG L, DONG J, GE J, et al. Efficient Frequency Domain-based Transformers for High-Quality Image Deblurring[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 5886-5895.
- [8] YANG F, ZHAI Q, LI X, et al. Uncertainty-guided transformer reasoning for camouflaged object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4146-4155.
- [9] ZHAO J X, LIU J J, FAN D P, et al. Egnnet: Edge guidance net-

- work for salient object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:8779-8788.
- [10] WANG S, XU Y X, ZENG D W, et al. Deep learning-based spectral reconstruction in camouflaged target detection[J]. International Journal of Applied Earth Observation and Geoinformation, 2024, 126: 103645.
- [11] DEEPTI Y, KUMAR M A, CHANDRA K T, et al. Detection and Identification of Camouflaged Targets using Hyperspectral and LiDAR data[J]. Defence Science Journal, 2018, 68(6): 540-540.
- [12] FAN D P, JI G P, ZHOU T, et al. Pranet: Parallel reverse attention network for polyp segmentation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention. 2020: 263-273.
- [13] MEI H, JI G P, WEI Z, et al. Camouflaged object segmentation with distraction mining [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8772-8781.
- [14] ZHAO Z, BAKAR A B E, RAZAK A B N, et al. Corrosion image classification method based on EfficientNetV2 [J]. Heliyon, 2024, 10(17): e36754.
- [15] ZHONG Y J, BO L, LYU T, et al. Detecting Camouflaged Object in Frequency Domain[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 4504-4513.
- [16] LIU M Z, DI X G. Extraordinary MHNet: Military high-level camouflage object detection network and dataset[J]. Neurocomputing, 2023, 549: 126446.
- [17] ZHAI Q, LI X, YANG F, et al. Mutual graph learning for camouflaged object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 12997-13007.
- [18] LYU Y Q, ZHANG J, DAI Y C, et al. Simultaneously localize, segment and rank the camouflaged objects[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 11591-11601.
- [19] LI A X, ZHANG J, LYU Y Q, et al. Uncertainty-aware joint salient object and camouflaged object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10071-10081.
- [20] FAN D P, JI G P, CHENG M M, et al. Concealed object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6024-6042.
- [21] HUANG Z, DAI H, XIANG T Z, et al. Feature Shrinkage Pyramid for Camouflaged Object Detection with Transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 5557-5566.
- [22] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: A new way to evaluate foreground maps[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 4548-4557.
- [23] JUTTEN J R, HARRISON J, KJOE M L R P, et al. A novel cognitive-functional composite measure to detect changes in early Alzheimer's disease: Test-retest reliability and feasibility[J]. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2018, 10: 153-160.
- [24] SINGHAL A, BEDI P. USteg-DSE: Universal quantitative Steganalysis framework using Densenet merged with Squeeze & Excitation net [J]. Signal Processing: Image Communication, 2024, 128: 117171-117171.
- [25] DICE L R. Measures of the amount of ecologic association between species[J]. Ecology, 2023, 26: 297-302.



LI Ang, born in 1992, research assistant. His main research interests include computer vision and AI security.



LIU Xunyun, born in 1989, associate research fellow. His main research interests include cloud computing and machine learning.

(责任编辑:柯颖)