

基于空间多尺度特征的伪造卫星图像检测方法

郝晓辰^{1,2,3}, 刘效勇^{2,3}

(1. 桂林理工大学 计算机科学与工程学院; 2. 桂林理工大学 物理与电子信息工程学院;
3. 广西嵌入式技术与智能系统重点实验室, 广西 桂林 541006)

摘要: 近年来, 深度伪造技术逐渐渗透到卫星成像领域, 对社会构成了由虚假卫星图像引发的严重威胁。为应对此类问题, 基于双色空间与空间多尺度特征融合提出了一种双流网络模型, 用于检测伪造卫星图像。具体而言, 设计了空间多尺度特征融合模块, 捕捉图像中的细微变化, 并结合SimAM注意力机制改进自校准卷积, 增强了模型对局部和全局信息的建模能力。此外, 引入双颜色空间, 在传统的RGB颜色空间的基础上, 增加了YCbCr颜色空间的特征, 进一步提升了模型在应对不同图像处理操作时的鲁棒性。实验结果表明, 该方法在UW数据集上达到了99.0%的F1分数, 优于现有检测方法。在常见的JPEG压缩、高斯噪声和模糊处理场景下的进一步测试表明, 该方法具有更强的鲁棒性, 证明了其在不同恶劣条件下的可靠性和实用性。

关键词: 深度伪造; 深度伪造地理; 自校准卷积; 多尺度特征融合; 颜色空间

DOI: 10.11907/rjdk.241635

开放科学(资源服务)标识码(OSID):

中图分类号: TP391.4

文献标识码: A

文章编号: 1672-7800(2025)009-0133-07



Detection Method of Forged Satellite Images Based on Spatial Multi-Scale Features

HAO Xiaochen^{1,2,3}, LIU Xiaoyong^{1,3}

(1. College of Physics and Electronic Information Engineering, Guilin University of Technology; 2. College of Computer Science and Engineering, Guilin University of Technology; 3. Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin 541006, China)

Abstract: In recent years, deepfake technology has increasingly infiltrated the field of satellite imaging, posing significant threats to society due to the potential dissemination of falsified satellite images. To address this issue, a dual-stream network model utilizing dual color spaces and spatial multi-scale feature fusion has been proposed for detecting forged satellite images. Specifically, a spatial multi-scale feature fusion module is designed to capture subtle variations within images. This module is further enhanced by incorporating the SimAM attention mechanism, which improves self-calibrated convolution, thereby enhancing the model's ability to capture both local and global information. Additionally, dual color spaces are utilized by integrating features from the YCbCr color space alongside the traditional RGB color space. This approach further enhances the model's robustness, particularly in handling various image processing operations. Experimental results demonstrate that the proposed method achieves an F1 score of 99.0% on the UW dataset, surpassing existing detection methods. Further testing under common conditions such as JPEG compression, Gaussian noise, and blurring shows that this method exhibits superior robustness, underscoring its reliability and practical utility in adverse scenarios.

Key Words: deepfake; deepfake geography; self-calibrated convolution; multi-scale feature fusion; color space

0 引言

随着人工智能技术的不断革新和硬件计算能力的显

著提升, 深度学习算法如今能够处理更加复杂、更大规模的数据集, 并可以实现更加繁杂的计算任务^[1]。这一进步极大地推动了人工智能生成内容(Artificial Intelligence Generated Content, 简称AIGC)的发展, 其中, 生成对抗网

收稿日期: 2024-07-29

扫描二维码阅读全文:



基金项目: 国家自然科学基金项目(62066011); 广西自然科学基金面上项目(2023GXNSFAA026057, 2022GXNSFAA035640)

作者简介: 郝晓辰(1995-), 男, CCF会员, 桂林理工大学计算机科学与工程学院硕士研究生, 研究方向为多媒体内容安全、图像取证;
刘效勇(1976-), 男, 博士, 桂林理工大学物理与电子信息学院教授、硕士生导师, 研究方向为信息安全、图像加密。本文通讯作者: 刘效勇。

络^[2](Generative Adversarial Networks, GANs)和扩散模型^[3](Denoising Diffusion Probabilistic Model, DDPM)等高性能图像生成方法,为高质量生成内容带来了无限的可能性。然而,AIGC技术的进步也催生了深度伪造技术(Deepfake)的迅速发展,对政治、经济和国家安全等方面构成了巨大威胁^[4,5]。例如,2022年,伪造的乌克兰总统泽连斯基投降视频被广泛传播,尽管最终被揭穿,但在短时间内已造成社会动荡。随着深度伪造现象的愈演愈烈,这些技术的不当使用已不再是简单的恶作剧或娱乐行为,而是逐渐演变成严重的犯罪行为。生成式人工智能的问题也逐渐转向风险问题^[6]。与此同时,深度伪造内容的主动取证也成为世界各国政府、企业及研究人员的关注热点^[7]。现如今,深度伪造技术逐渐渗透到卫星成像领域,产生了“深度伪造地理(Deepfake geography)”的新问题^[8]。卫星图像广泛应用于环境监测、灾害预警和国防安全等诸多领域,而伪造的卫星图像可能被用于散布虚假信息、实施非法活动或干扰军事决策,从而带来巨大的安全隐患^[9-11]。

1 相关研究

AIGC生成的虚假卫星图像非常逼真,几乎无法用肉眼区分。尽管目前在深度伪造检测方面取得了一些进展,但由于缺乏相关数据集,检测深度伪造卫星图像的研究仍然有限。为此,Zhao等^[8]提出了一种基于CycleGAN^[12]生成假卫星图像的方法,创建了第一个深度伪造卫星图像UW数据集。在该数据集中,伪造的卫星图像基于CycleGAN生成,数据来源为塔科马(Tacoma)、西雅图和北京三个城市的卫星图像;真实卫星图像则从谷歌地球的卫星图像中采集。部分UW数据集的样本如图1所示。

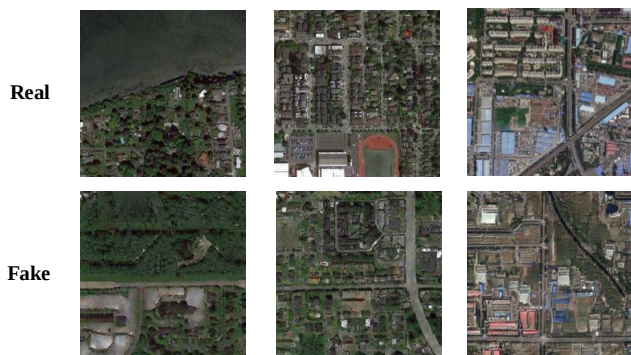


Fig. 1 Some samples from the UW dataset

图1 UW数据集中的部分样本

与此同时,基于早期研究中发现的GANs生成的图像在颜色、纹理、细节等视觉特征以及频域特征上与真实图像存在差异的理论,该团队通过独立t检验,从颜色空间、直方图和频率3个类别中选取了26个手工制作的特征,用于识别真实和伪造卫星图像之间显著的平均差异。这些特征随后被输入支持向量机(Support Vector Machine, SVM)进行虚假卫星图像的检测。然而,基于传统手工提

取特征的检测方法在面对日益复杂和多样的伪造手段时,往往难以应对^[13]。为此,Fazza等^[14]尝试使用卷积神经网络(Convolutional Neural Network, CNN)检测伪造卫星图像,并利用在ImageNet数据集上预训练的四个CNN模型:ResNet50^[15]、VGG16^[16]、InceptionV3^[17]和Xception^[18]开展实验。通过实验,他们评估了这些模型在伪造卫星图像检测中的性能和可行性,但是检测精度仍需进一步提升。

为了应对深度伪造地理的威胁,开发准确有效的伪造卫星图像检测方法已成为亟待解决的问题。为提供有力的技术支持,本文改进了原始的自校准卷积,设计了空间多尺度自校准卷积块,并进一步搭建了空间多尺度自校准卷积神经网络。结合该网络基于双色空间设计的双流检测模型,通过研究和实验验证,本文方法在检测精度和鲁棒性方面展现了卓越的性能,能够有效应对这一日益严峻的挑战。

2 相关技术

2.1 自校准卷积神经网络

下游任务的分类性能依赖于上游卷积网络的特征提取能力。自校准卷积神经网络(Self-Calibration Net, SC-Net)在不调整整个网络模型架构的基础上,通过改进基本的卷积模块提升网络模型的性能^[19]。基于残差结构,SC-Net设计了一个即插即用的自校准卷积模块(Self-Calibration Module, SCmodule),替代了传统的卷积块。与传统的2D卷积不同,SC模块将输入数据均匀划分并分别传入2个分支。上分支为自校准分支,对传入数据进行均值下采样、卷积特征变换、双线性下采样以及空域注意力特征图融合等操作。由于自校准分支仅考虑空域位置信息,避免了不干扰信息的影响,因此可以极大提升输出特征的感受野。另一部分输入数据传入下分支,下分支为常规卷积变换分支,保留数据的背景关系及上下文。最终,将处理后的特征进行拼接,作为模型的输出数据。SC模块每个空间位置的感受野不受预定义的卷积核大小控制,在不引入额外参数的情况下,能够帮助模型更完整、更准确地提取特征,同时增强特征的可区分。基于以上优势,本文选用SCmodule作为基准,并对其进行改进。

2.2 3D注意力机制SimAM

在深度学习和计算机视觉领域,注意力机制借鉴了人类大脑的注意力原理,通过增强模型对关键特征的关注重来提升模型性能。本文采用了一种新颖高效的3D注意力模块SimAM,该模块无需额外参数,具备即插即用的轻量级特性,旨在强化卷积神经网络中特征的表达能力^[20]。与1D注意力平等对待一个通道内所有神经元、2D注意力平等对待同一空间位置上所有神经元不同,3D注意力采用统一权值,分别评估每个神经元的重要性。这种方法有效帮助注意力机制学习跨通道和跨空间的变化权重,具有更

强的灵活性。SimAM通过能量函数来挖掘神经元的重要性,它能够任何中间特征张量作为输入并进行相应的转换。输出张量与输入张量具有相同尺寸,但是表征得到了增强,SimAM的结构示意图如图2所示。

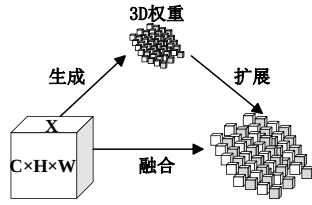


Fig. 2 3D attention SimAM structure diagram

图2 3D注意力SimAM结构图

神经元科学研究发现,在图像处理任务中,应该给予表现出明显空间抑制效果的神经元更高的优先级。找到这些神经元的最简单实现,是测量一个目标神经元与其他神经元之间的线性可分性。

基于以上理论,SimAM注意力的实现需要估计单个神经元的重要性,因此SimAM注意力为每个神经元定义了能量函数,其中,初始能量函数如公式(1)所示。

$$e_i(w, b, e, x_i) = (y_i - \hat{t})^2 \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (1)$$

假设,输入特征图为 $X \in R^{H \times C \times W}$ (H, C, W 为输入数据的维度),能量公式中, M 代表每个通道上的能量个数, t 和 x_i 是输入特征图单通道上的目标神经元和其他神经元。 $\hat{t} = w_t t + b_t$ 和 $\hat{x}_i = w_i x_i + b_i$, 是 t 和 x_i 的线性变换,其中, w_i 、 b_i 分别代表线性变换的权重和偏置, i 是空间维度上的索引。为了简便衡量当前神经元和其它神经元之间的差异,将当前神经元设为正类,其余设为负类,采用二值标签, $y_t = 1, y_0 = -1$, 正则变换。最终能量函数如公式(2)所示。

$$e_i(w, b, e, x_i) = (1 - w_t t - b_t)^2 + \lambda w_t^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - w_i x_i - b_i)^2 \quad (2)$$

3 模型构建

3.1 空间多尺度特征融合模块 SMFM

本文设计的空间多尺度特征融合模块(Spatial Multi-Scale Feature Fusion Module, SMFM),旨在实现调制多尺度特征融合。该算法通过聚合不同尺寸卷积核提取的特征来捕捉不同尺度的上下文信息,增强模型的表达能力。SMFM由2个部分构成:多尺度特征聚合部分和特征图调制部分,结构如图3所示。

多尺度特征聚合部分使用 1×1 、 3×3 和 5×5 三种不同大小的卷积核,对输入特征图进行卷积。 1×1 卷积主要用于调整通道数和捕捉局部特征; 3×3 卷积提供中等范围的感受野,平衡局部与全局信息; 5×5 卷积则提供更广的感受野,捕捉更大范围的上下文信息。在每个卷积操作之后,都会连接一个批量归一化层(Batch Normalization Layer, BN层),并通过 ReLU 激活函数进行处理。这一结构不仅有助

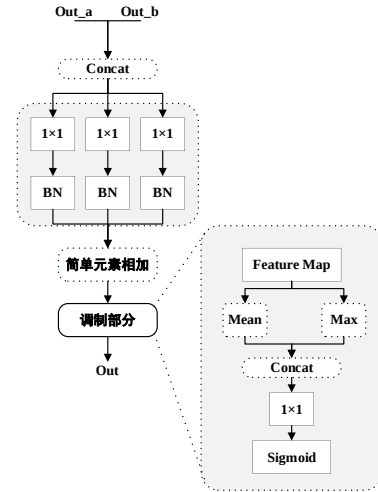


Fig. 3 Structure diagram of SMFM

图3 SMFM结构图

于稳定训练过程,还能有效提升模型的收敛速度。原理如公式(3)所示。

$$O_k = \text{ReLU}(\text{BN}(C_{k \times k}(x))) \quad (3)$$

其中, k 取值为 1、3、5, 分别代表不同尺寸的卷积核。聚合这 3 种尺度的输出,使用简单的逐元素加法将 1×1 、 3×3 和 5×5 卷积后的特征图相加。这种直接相加的方法相较于拼接简单但有效,合并不同尺度的上下文信息,融合来自不同尺度的特征,使得最终的特征图既包含细节信息也包含较大范围的上下文信息,聚合原理如公式(4)所示。

$$F(x) = \text{ReLU}(\text{BN}(C_{1 \times 1}(x)) + \text{BN}(C_{3 \times 3}(x)) + \text{BN}(C_{5 \times 5}(x))) \quad (4)$$

特征图调制部分的设计受空间注意力机制的启发,具体如下:首先,计算输入特征图的通道平均值和最大值,分别提取全局信息和突出特征;其次,将通道平均值和最大值合并成一个双通道的特征图,综合两种统计信息,为后续处理步骤提供更丰富的上下文;其次,通过一个 1×1 的卷积核对合并后的双通道特征图进行卷积,并调制特征图在每个位置的激活值,使模型能够动态地聚焦于每个位置上最重要的特征;最后,使用 Sigmoid 函数来激活输出,将卷积结果转换为一个权重图用于调制原始输入特征图,调节每个空间位置的特征强度,突出重要的区域并抑制不重要的区域,从而使模型更加关注于对最终任务有用的信息,调制原理如公式(5)所示。

$$A(x) = \text{Sigmoid}(C_{1 \times 1}(\text{Concat}(\text{AvgPool}(x), \text{MaxPool}(x)))) \quad (5)$$

其中,模块内部先计算特征图 $F(x)$ 的通道平均值 $\text{AvgPool}(x)$ 和最大值 $\text{MaxPool}(x)$,合并两个通道的特征图后进行 1×1 卷积和 Sigmoid 函数激活,经过调制处理后,输出权重图 $A(x)$ 。

调制模块的权重图 $A(x)$ 与融合模块输出的特征图 $F(x)$ 进行逐元素乘法 $\odot$, 确保 SMFM 的输出特征图在空间上得到了有效的重标定,产生的新特征图 $F'(x)$ 如公式(6)所示。

$$F'(x) = A(x) \odot F(x) \quad (6)$$

综上所述,SMFM通过聚合多尺度特征并调制特征图,增强了模型的感受野及捕捉不同尺寸细节的能力,从而提高了模型的性能和泛化能力。

3.2 空间多尺度自校准卷积

本文在SCmodule的基础上进行了改进。首先,在2个分支的特征向量拼接之前,于每个分支的末尾分别添加了一层SimAM注意力机制,以过滤无关信息并集中关注有效信息;其次,将过滤后的特征向量进行拼接,并使用SMFM模块进一步处理,以增强模型的感受野和上下文信息聚合能力;最后,输出处理结果。改进后的空间多尺度自校准卷积块(Spatial Multi-Scale Feature Self-Calibration Module, SSCmodule)结构如图4所示。

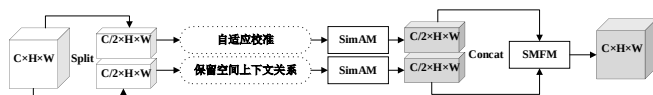


Fig. 4 Spatial multi-scale self-calibrated convolution module

图4 空间多尺度自校准卷积 SSCmodule

3.3 数据预处理

颜色空间在图像处理领域中对特征提取的效果具有显著影响。清华大学研究团队利用7种不同的颜色空间,分别传入7个单独的DenseNet,最终通过Denselayer进行分类,取得了显著的性能提升^[21]。这表明,颜色空间(即原始RGB图像的变换)能够显著影响分类准确性。Chen等^[22]在前述研究的基础上,通过在不同颜色空间计算从原始图像中提取的特征与从相应后处理图像中提取的特征之间的欧氏距离,并进行相似性分析,发现颜色空间的选择直接影响特征提取的效率和模型的性能。实验结果表明,不同的颜色空间能够从不同角度描述图像内容。

某些颜色空间对环境变化具有较强的抵抗性,适用于在复杂背景或光照条件变化下的图像分类任务,从而提升模型的鲁棒性。并且,同一颜色空间中的色度分量与亮度分量在抵抗不同后处理操作时表现不同,这一线索也有助于改进分类结果,提升模型在分类任务中的性能。尽管中值滤波等增强方法可以过滤噪声等干扰信息,但由于检测图像的后期处理方法往往是未知的,不同的增强策略在不同类别的数据集上表现各异,难以针对性地使用增强方案。因此,受前述研究启发,考虑到从原始图像中提取的特征与从相应后处理图像中提取的特征之间相似度越强,模型的鲁棒性越高,且YCbCr颜色空间具有均匀颜色分布并能够有效分离亮度和色度特征的优势,本文采用受检图像的RGB空间表示与YCbCr空间表示作为双通道输入数据,以提升模型在实际应用中的鲁棒性。

3.4 双流伪造卫星图像检测模型

检测模型由基于SSCmodule空间多尺度自校准卷积重新设计的空间多尺度自校准网络(Spatial Multi-Scale Feature Self-Calibration Net, SSCNet)作为特征提取通道。每个检测通道采用4个含有不同数量SSCmodule的SSCList

层,以更好地提取高级语义信息,预防梯度消失。在1个独立的SSCList层当中,每个SSCmodule采用相同的网络结构设计和尺寸,不同的SSCList层内SSCModule数量不同。首先,输入特征通道的图像通过Transfer.Resize函数转化成尺寸为64×64的图像,传入1个包含64个尺寸为7×7、步长为2的过滤器进行提取。提取表示后通过最大池化层(Max_Pool)进行下采样。为了匹配输入数据的维度并优化池化层的采样效果,将Max_Pool的步长调整为1。其次,依次通过以下模块提取特征表示:包含1个SSCmodule子模块的SCList1、包含1个SSCmodule子模块的SCList2、包含2个SSCmodule子模块的SCList3以及包含1个SSCmodule子模块的SCList4。采用1:1:2:1的层数设计,旨在保证特征提取能力的同时,避免参数量呈指数级增长而导致网络参数冗余和效率下降。然后,经过自校准池化(AdaptiveAvgPool)操作输出维度为2048×1×1的特征张量。SSCNet使用Batch_Norm稳定网络,防止过拟合,加速收敛,并提升网络性能。最后,在每个卷积层使用Relu激活函数对输出进行非线性化处理。

在2个通道AdaptiveAvgPool操作之后将2个特征张量合并馈入全连接层。经过全连接合并后的特征张量经过2次卷积与反卷积操作,维度由2048×2依次转变为1024、2×1。最终,通过Softmax函数对全连接层的输出数据进行归一化,得到检测图像的结果分类。本文方法的完整结构如图5所示。

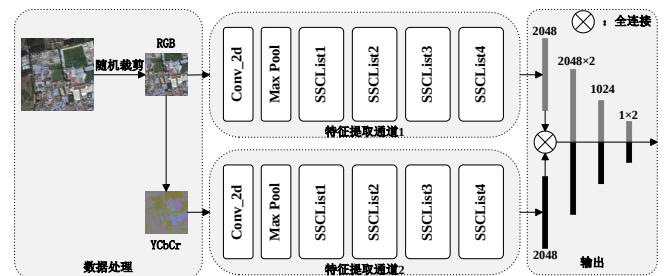


Fig. 5 Dual-stream forged satellite image detection model

图5 双流伪造卫星图像检测模型

4 实验与结果分析

为验证所提方法的有效性,本研究将其与现有方法进行了全面对比分析。对比方法主要包括以下两类:①Zhao等^[8]提出的基于手工特征的方法;②基于迁移学习的深度卷积神经网络方法,包括VGG16、InceptionV3和Xception等经典模型。此外,为了评估模型在处理不同后处理图像时的鲁棒性,对模型在模糊、JPEG压缩和高斯噪声3种不同数据集上的性能进行进一步的实验。

4.1 实验环境设置

实验配置如下:CPU选用Intel(R)Silver 4210R@2.40 GHz 2.39 GHz双处理器,GPU选用RTX3060 12 GB显卡。训练测试平台为Python3.8,使用Pytorch的深度学习库。

实验设置如下:在训练和验证阶段,学习率设定为 0.001,批处理大小为 32;在测试阶段,批处理大小为 64。测试进行 20 轮,最终结果取平均值,并保留至小数点后三位。本文使用 Zhao 等^[8]公开的 UW 数据集进行实验验证。该数据集共包含 8 064 张卫星图像,其中 4 032 张为 256×256 像素的真实卫星彩色图像,另外 4 032 张为 256×256 像素的伪造卫星彩色图像,这些伪造图像由 CycleGAN 生成。实验将 UW 数据集按照 9:1 的比例分别用于训练和测试。

4.2 评价指标

实验采用图像二分类任务中广泛使用的 3 种性能指标对模型进行评估,具体包括精度(Precision, P)、召回率(Recall, R)和 F1 分数(F1-score, F1)。这 3 种评价指标的定义如式(7)所示。

$$\begin{cases} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \\ F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \end{cases} \quad (7)$$

其中,真阳性(true positive, TP)表示被正确分类为真实卫星图像的真实卫星图像数量;假阳性(false positive, FP)表示被错误分类为真实卫星图像的伪造卫星图像数量;假阴性(false negative, FN)表示被错误分类为伪造卫星图像的真实卫星图像数量。

4.3 对比实验

4.3.1 原始数据集对比实验

在未经处理过的原始 UW 测试集上的对比实验结果如表 1 所示。

Table 1 Experimental results on the original dataset

表 1 原始数据集实验结果

方法	P	R	F1
Zhao ^[8]	0.827	0.919	0.871
InceptionV3 TL ^[17]	0.927	0.926	0.926
VGG16 TL ^[16]	0.965	0.965	0.965
Xception TL ^[18]	0.953	0.953	0.953
Ours	0.992	0.987	0.990

实验结果表明,本文方法在伪造卫星图像检测任务中显著优于现有方法,在所有评估指标上均取得了最高的分数。因手工提取特征的方法受限于特征提取的效果和模型的复杂度,在捕捉伪造卫星图像特征方面的能力有限,所以 Zhao 等^[8]提出的方法表现相对较差。而基于迁移学习的 CNN 表现强于基于手工特征的方法。相比 Zhao 等^[8]的方法以及基于迁移学习的 InceptionV3、VGG16 和 Xception 模型,本文设计的双流检测模型在特征提取和分类能力方面聚合了更全面的上下文信息,表现最为突出。

4.3.2 JPEG 压缩数据集对比实验

在图像传输过程中,数据通常会经过 JPEG 压缩处理。JPEG 压缩通过调节质量因子(Quality Factor, QF)来权衡图像质量与编码比特率之间的关系,从而实现了对图像压缩过

程中图像质量与文件大小之间平衡的控制。QF 的取值范围通常为 1~100 之间的整数,QF 数值越高,压缩比越低,压缩后图像的质量就越高,文件大小越大;QF 数值越低,压缩比越高,压缩后图像质量就会明显下降,出现更多失真,且文件大小更小。为了评估压缩失真影响下模型的鲁棒性,分别使用 QF 值为 75、85 和 95 对测试图像进行 JPEG 压缩,并设计了 3 组对比实验对模型进行测试,实验结果如表 2 所示。

Table 2 Experimental results on the JPEG compressed dataset

表 2 JPEG 压缩数据集实验结果

因子	方法	P	R	F1
75	Zhao ^[8]	0.884	0.920	0.902
	InceptionV3 TL ^[17]	0.892	0.880	0.879
	VGG16 TL ^[16]	0.958	0.958	0.958
	Xception TL ^[18]	0.913	0.911	0.911
	Ours	0.970	0.975	0.972
85	Zhao ^[8]	0.858	0.936	0.895
	InceptionV3 TL ^[17]	0.905	0.900	0.899
	VGG16 TL ^[16]	0.960	0.960	0.960
	Xception TL ^[18]	0.941	0.941	0.941
	Ours	0.972	0.982	0.977
95	Zhao ^[8]	0.850	0.889	0.869
	InceptionV3 TL ^[17]	0.924	0.923	0.923
	VGG16 TL ^[16]	0.962	0.962	0.962
	Xception TL ^[18]	0.946	0.946	0.946
	Ours	0.989	0.982	0.986

通过表 2 中对比可以看出,本文所提方法在不同压缩质量因子下均表现出色,展示了卓越的鲁棒性。尽管基于手工特征的方法利用图像质量测量特征和频率特征可以捕获压缩图像中的更多判别信息,提升了 Zhao 等^[8]的检测指标,但本文所提方法依然表现出最佳的检测性能。无论是在较低还是较高的压缩质量因子下,均能够保持高 Precision 和 Recall,且 F1 得分始终保持在最高水平。相较于其他方法中使用的单一颜色空间数据,本方法通过引入双颜色空间数据预处理,增强了模型的鲁棒性和泛化能力,有效地捕捉了更多伪造图像的细节特征。特别是在高压压缩质量因子下,依然能够保持接近完美的检测能力。这表明本文方法不仅在不同压缩质量下具有出色的检测性能,还具有很强的适应性和鲁棒性。

4.3.3 高斯噪声数据集对比实验

卫星图像在采集和传输过程中,由于设备信道的缘故,容易受到高斯噪声的影响。高斯噪声的强度由噪声因子进行量化和描述,通常定义为高斯分布中噪声的标准差。噪声因子具体表示噪声信号相对于理想信号的扰动程度,定义了噪声在平均值附近的散布,噪声值与噪声因子的关系如式(8)所示。

$$n(x, y) = \sigma \cdot \text{randn}(x, y) + \mu \quad (8)$$

其中, $n(x, y)$ 是图像中的噪声值,标准差 σ 是噪声因子, $\text{randn}(x, y)$ 是服从均值为 0,标准差为 1 的标准正态分布的随机变量,通过与噪声因子相乘控制噪声幅度。高斯

噪声对图像高频特征的影响极大,对检测模型具有强烈的攻击性。为评估噪声影响下的模型性能水平,使用标准差为 0.02 的噪声因子的测试集进行实验,实验结果如表 3 所示。

Table 3 Experimental results on the Gaussian noise dataset

表 3 高斯噪声数据集实验结果

因子	方法	P	R	F1
0.02	Zhao ^[8]	0.814	0.851	0.832
	InceptionV3 TL ^[17]	0.895	0.887	0.886
	VGG16 TL ^[16]	0.948	0.946	0.946
	Xception TL ^[18]	0.884	0.865	0.863
	Ours	0.963	0.980	0.971

实验结果表明,在添加噪声的情况下,本文方法相较于其他对比方法,依然表现出最高的检测性能,且显著优于其他方法,证明了其在噪声环境中具有较强的鲁棒性。

4.3.4 高斯模糊数据集对比实验

除采集和传输过程外,由于外部环境条件的不稳定性(如卫星运动、采集设备移动等因素),卫星图像常会出现模糊现象。这种模糊化处理会导致数据极易丢失重要的特征点,从而影响检测模型的性能。在高斯模糊处理中,窗口是指用于模糊操作的像素区域或内核的大小和形状。窗口的大小决定了高斯模糊在图像中的局部应用范围,对模糊效果的强度和范围有直接影响。通常,窗口大小以奇数表示,以确保模糊处理时始终有一个明确的中心像素。随着窗口尺寸的增大,参与模糊计算的像素数量增加,使得高斯模糊的效果更加显著,从而导致图像变得更加模糊。为评估模糊处理场景下模型的鲁棒能力,采用 5×5 与 9×9 两种窗口尺寸的高斯模糊技术对测试集进行处理,并展开实验,实验结果如表 4 所示。

Table 4 Experimental results on the Gaussian blur dataset

表 4 高斯模糊数据集实验结果

窗口	方法	P	R	F1
5×5	InceptionV3 TL ^[17]	0.876	0.865	0.864
	VGG16 TL ^[16]	0.911	0.958	0.901
	Xception TL ^[18]	0.880	0.856	0.853
	Ours	0.988	0.845	0.910
9×9	InceptionV3 TL ^[17]	0.815	0.772	0.764
	VGG16 TL ^[16]	0.883	0.867	0.866
	Xception TL ^[18]	0.816	0.711	0.698
	Ours	0.983	0.757	0.855

在 5×5 窗口模糊处理下,本文方法的 Precision 显著优于其他对比方法,尽管 Recall 值(0.845)相对较低,但本文方法仍取得了最高的 F1 得分;在 9×9 窗口模糊处理下,本文方法的 Precision 值(0.983)仍然保持在较高水平,但 Recall 相对较低,且 F1 得分为 0.855。实验结果表明,本文方法在模糊处理条件下能够保持较高的检测精度,但在较大窗口模糊处理时召回率出现一定程度的下降。尽管如此,本文方法的 F1 分数仍优于其他对比方法,保持了较优的综合性能。

4.4 消融实验

为证明设计的合理性及有效性,本文设置了 2 个消融实验:组件消融实验与通道消融实验。

4.4.1 组件消融实验

组件消融实验评估不同组件对检测性能的影响,通过控制变量对模型进行消融实验,结果如表 5 所示。

Table 5 Ablation study results

表 5 消融实验结果

架构	P	R	F1
SC+SMFM+SimAM	0.992	0.987	0.990
SC+SimAM	0.989	0.985	0.987
SC	0.985	0.982	0.983

实验结果表明,结合 SC、SMFM 和 SimAM 的完整架构(SC+SMFM+SimAM)性能最佳,Precision、Recall 和 F1 得分分别为 0.992、0.987 和 0.990。去除 SMFM 后(SC+SimAM),性能略有下降,但仍然表现优异,F1 得分为 0.987。仅使用 SC 时,F1 得分为 0.983。实验结果证明了 SMFM 和 SimAM 组件在提升模型性能方面具有重要作用。

4.4.2 通道消融实验

通道消融实验评估双流网络设计的合理性,通过控制变量对模型进行消融实验,分别评估 RGB 流、YCbCr 流与双流网络的性能,结果如表 6 所示。

Table 6 Ablation study results

表 6 消融实验结果

通道	P	R	F1
RGB+YCbCr	0.992	0.987	0.990
RGB	0.985	0.983	0.984
YCbCr	0.981	0.980	0.981

实验结果表明,结合 RGB 和 YCbCr 通道的双流网络(RGB+YCbCr)性能最佳,Precision、Recall 和 F1 得分分别为 0.992、0.987 和 0.990。使用 RGB 流,性能略有下降,F1 得分为 0.984。仅使用 YCbCr 通道时,性能最低,F1 得分为 0.981。实验结果证明了本文方法通过结合 RGB 和 YCbCr 通道对模型性能提升具有显著贡献,同时验证了双流网络设计相较于单流网络具有明显的优越性。

5 结语

本文针对深度伪造地理问题,设计了一种双流检测模型,用于区分真实与伪造的卫星图像。提出了一种空间多尺度特征融合模块,并结合 3D 注意力机制改进了自校准卷积模块,将其命名为空间多尺度自校准卷积块。本文方法基于双色空间作为数据输入,使用空间多尺度自校准卷积块,重新设计特征提取通道的骨干网络。实验结果表明,与现有检测方法相比,本文方法不仅展现出更优的检测性能,而且对不同类型的后处理操作也表现出更强的鲁棒性优势。如今,伪造卫星图像检测仍然处于初级阶段,目前的 UW 数据集也仅包括 3 个城市,具有一定的局限性。

在未来研究中,将重点扩展更多类型的数据集,涵盖更多城市,并利用高性能的GANs算法与扩散模型生成更加逼真的伪造卫星图像,以进一步完善数据集。同时,结合具有更强全局特征提取能力的技术,设计更加实用的伪造卫星图像检测算法。

参考文献:

- [1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.
- [2] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [3] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33:6840-6851.
- [4] CHADHA A, KUMAR V, KASHYAP S, et al. Deepfake: an overview [C]// Proceedings of Second International Conference on Computing, Communications, and Cyber-Security, 2021:557-566.
- [5] WANG R Y, CHU B L, YANG Z, et al. An overview of visual DeepFake detection techniques[J]. Journal of Image and Graphics, 2022, 27(1): 43-62.
王任颖, 储贝林, 杨震, 等. 视觉深度伪造检测技术综述[J]. 中国图象图形学报, 2022, 27(1):43-62.
- [6] DING X D. Legal regulation of artificial intelligence risks——an example from the EU artificial intelligence act[J]. Science of Law (Journal of Northwest University of Political Science and Law), 2024, 42(5):3-18.
丁晓东. 人工智能风险的法律规制——以欧盟《人工智能法》为例[J]. 法律科学(西北政法大学学报), 2024, 42(5):3-18.
- [7] LIANG R G, LYU P Z, ZHAO Y, et al. A survey on audiovisual deepfake detection technologies[J]. Journal of Information Security, 2020, 5(2):1-17.
梁瑞刚, 吕培卓, 赵月, 等. 视听深度伪造检测技术研究综述[J]. 信息安全学报, 2020, 5(2):1-17.
- [8] ZHAO B, ZHANG S, XU C, et al. Deep fake geography? when geospatial data encounter artificial intelligence[J]. Cartography and Geographic Information Science, 2021, 48(4):338-352.
- [9] DU S H, LI W, XING J H, et al. Change detection of open-pit mines based on FM-UNet++ and GF-2 satellite images [J]. Coal Geology & Exploration, 2023, 51(7):130-139.
杜守航, 李炜, 邢江河, 等. 基于 FM-UNet++ 和高分二号卫星影像的露天矿区范围变化检测[J]. 煤田地质与勘探, 2023, 51(7):130-139.
- [10] FU Q, LUO W L, LYU J X. Land utilization change detection of satellite remote sensing image based on AlexNet and support vector machine[J]. Laser & Optoelectronics Progress, 2020, 57(17):282-290.
付青, 罗文浪, 吕敬祥. 基于 AlexNet 和支持向量机相结合的卫星遥感影像土地利用变化检测[J]. 激光与光电子学进展, 2020, 57(17):282-290.
- [11] MA X, XU J, PAN J, et al. Detection of marine oil spills from radar satellite images for the coastal ecological risk assessment [J]. Journal of Environmental Management, 2023, 325:116637.
- [12] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Venice: 2017 IEEE International Conference on Computer Vision (ICCV), 2017.
- [13] LI X R, JI S L, WU C M, et al. Survey on deepfakes and detection techniques[J]. Journal of Software, 2021, 32(2):496-518.
李旭嵘, 纪守领, 吴春明, 等. 深度伪造与检测技术综述[J]. 软件学报, 2021, 32(2):496-518.
- [14] FEZZA S A, OUIS M Y, KADDAR B, et al. Evaluation of pre-trained cnn models for geographic fake image detection [C]// 2022 IEEE 24th International Workshop on Multimedia Signal Processing, 2022:1-6.
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. <https://arxiv.org/abs/1409.1556>.
- [17] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]// Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [18] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]// Honolulu: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [19] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions [C]// Seattle: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [20] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks [C]// Proceedings of the 38th International Conference on Machine Learning, 2021:11863-11874.
- [21] GOWDA S N, YUAN C. ColorNet: investigating the importance of color spaces for image classification [C]// Computer Vision-ACCV 2018: 14th Asian Conference on Computer Vision, 2019:581-596.
- [22] CHEN B, LIU X, ZHENG Y, et al. A robust GAN-generated face detection method based on dual-color spaces and an improved Xception [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(6):3527-3538.

(责任编辑:陈维捷)