计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.009892] [CSTR: 32024.14.csa.009892]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于多尺度感知学习的图像篡改检测与定位①

徐　悦[1,2], 袁程胜[1,2], 刘庆程[1,2], 夏志华[3]

[1](南京信息工程大学 计算机学院、网络空间安全学院, 南京 210044)
[2](南京信息工程大学 数字取证教育部工程研究中心, 南京 210044)
[3](暨南大学 网络空间安全学院, 广州 510632)
通信作者: 袁程胜, E-mail: yuancs@nuist.edu.cn

摘　要: 为了解决现有图像篡改检测方法在检测定位性能与鲁棒性方面的不足, 本文提出了一种多尺度感知学习网络 (MsPL-Net). 首先, 为了扩展感受野并解决图像后处理和操作类型多样导致的鲁棒性弱的难题, 提出了一种分层密集链接多尺度扩展卷积模块 (MSDCM). 该模块可放大感受野以捕捉多尺度特征信息, 同时保持输入图像的高分辨率表示, 无缝提取复杂的图像细节和边缘信息. 其次, 为了解决篡改大小敏感性导致的篡改边缘位置模糊问题, 提出了一种由全局注意力、局部注意力和门控特征调节器组成的信息互补感知注意力模块 (ICPAM). 全局注意可以捕捉图像的整体形状、结构或背景信息, 而局部注意可以学习图像的局部区域和具体细节, 两者交互融合, 提高定位精度. 门控特征调节器采用精细嵌入从全局和局部特征图中过滤出不相关的特征和噪声响应, 引导下游识别和学习由不同篡改技术引起的异常纹理、边缘变化和其他特征信息. 最后, 设计一种新的联合损失函数, 进一步提高网络的检测性能和定位准确率. 相较于最新工作, 本文方法的检测准确率提高了 2.3%. 此外, 在鲁棒性和泛化性上同样表现出较好的性能, 以及篡改区域定位更精确和清晰.

关键词: 图像篡改检测; 鲁棒性; 图像后处理; 多尺度扩张卷积; 信息互补感知注意力

引用格式: 徐悦,袁程胜,刘庆程,夏志华.基于多尺度感知学习的图像篡改检测与定位.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9892.html

Multi-scale Perceptual Learning for Image Manipulation Detection and Localization

XU Yue[1,2], YUAN Cheng-Sheng[1,2], LIU Qing-Cheng[1,2], XIA Zhi-Hua[3]

[1](School of Computer Science, School of Cyber Science and Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)
[2](Engineering Research Center of Digital Forensics of Ministry of Education, Nanjing University of Information Science & Technology, Nanjing 210044, China)
[3](College of Cyber Security, Jinan University, Guangzhou 510632, China)

Abstract: To address the shortcomings of existing image tampering detection methods in terms of detection and localization performance as well as robustness, a multi-scale perceptual learning network (MsPL-Net) is proposed. Firstly, to expand the receptive field and address the issue of weak feature robustness resulting from diverse image post-processing and operation types, a hierarchical dense linked multi-scale dilated convolution module (MSDCM) is introduced. This module expands the receptive field to capture multi-scale feature information while preserving the high-resolution representation of input images, seamlessly extracting intricate image details and edge information. Secondly, to solve the problem of blurred tampered edge positions caused by sensitivity to tampering size, an information complementary perception attention module (ICPAM) is proposed, consisting of global attention, local attention, and a gated feature modulator. The global and local attention mechanisms operate in parallel and complement each other:

① 基金项目: 国家自然科学基金 (62102189, 62122032); 国家社会科学基金 (2022-SKJJ-C-082)
收稿时间: 2024-12-04; 修改时间: 2024-12-25; 采用时间: 2025-01-24; csa 在线出版时间: 2025-04-25

through feature interaction and fusion, the model's representational capacity is enhanced, leading to improved localization performance. Global attention captures the overall shape, structure, or background information of the image, while local attention focuses on learning the local regions and specific details of the image. The two mechanisms interact and integrate to enhance positioning accuracy. The gated feature modulator employs fine embeddings to filter out irrelevant features and noise responses from the global and local feature maps. This facilitates downstream recognition and learning of abnormal textures, edge changes, and other feature information caused by different tampering techniques. Finally, a novel joint loss function is designed to further enhance the detection performance and localization accuracy of the network. Compared with the latest works, the detection accuracy of the proposed method is improved by 2.3%. In addition, the proposed method demonstrates excellent performance in terms of robustness and generalization, offering more accurate and clear localization.

Key words: image tampering detection; robustness; image post-processing; multi-scale dilated convolution; information complementary perceptual attention

人工智能和大模型的兴起为人们提供了展示才能的舞台, 而图像编辑软件的发展则极大地简化了图像编辑操作. 与此同时, 负面安全事件也层出不穷. 恶意攻击者滥用图像编辑技术, 企图欺骗公众, 导致图片篡改相关的安全问题频发. 当下, 许多高质量的篡改图像在视觉上几乎难以察觉任何篡改的痕迹. 因此, 解决篡改图像带来的安全性威胁是当下迫切需要解决的问题, 图像篡改检测与定位方法必须不断发展以应对新形势下日益复杂的图像篡改挑战.

图像篡改方法大致可分为语义相关和语义无关的两类, 区别主要在于篡改对图像语义信息的影响程度. 语义相关的篡改方法显著改变图像的语义内容, 主要包括复制粘贴, 移动删除和拼接组合等操作. 而语义无关的图像篡改方法则主要涉及旋转、缩放、模糊等操作, 这些操作对图像语义的影响较为轻微, 主要用于修饰篡改后的图像. 然而, 当这两类方法被联合使用时, 它们所产生的篡改痕迹会相互交织, 从而显著提升检测的难度. 因此, 在设计图像篡改检测方法时, 必须同时考虑语义相关和语义无关这两种篡改类型.

近年来, 越来越多的图像篡改检测与定位方法不断涌现, 但这些方法仍面临若干显著挑战. 第一, 鲁棒性弱: 篡改图片往往经过语义篡改后附加后处理操作以掩盖痕迹, 导致后处理与语义篡改的痕迹相互交织, 提取的特征混杂着篡改伪影与后处理噪声, 极大地增加了检测难度. 在某些情况下, 后处理操作甚至可能引发误检或漏检, 进而削弱整体检测效能. 第二, 篡改区域尺度的多样性对检测适应性构成挑战: 由于图像篡改的对象和篡改区域的形状、大小、位置都是随机的, 当前方法在应对不同尺度篡改区域时表现各异. 部分网络模型在常规尺度下表现出色, 但在极端尺度下却可能出现漏检或误检的问题. 第三, 定位篡改区域边缘模糊问题: 篡改操作遗留的伪影使得特征图的局部特征上下文关系及细节更为复杂, 简单的特征融合方法难以捕捉关键信息, 甚至可能导致篡改区域边缘信息错位, 进而造成定位模糊或定位错误. 为此, 本文的主要贡献如下.

(1) 针对图像后处理多样性和操作类型复杂导致的特征鲁棒性不足问题, 本文设计了一个多尺度扩张卷积模块 (MSDCM) 用于特征提取. 它可以放大感受野以捕捉多尺度特征信息, 同时保持输入图像的高分辨率表示, 从而平滑且连续地提取出复杂的图像细节和边缘信息.

(2) 为解决篡改大小敏感性引发的篡改边缘位置模糊难题, 本文提出了一种由全局注意力、局部注意力和门控特征调节器组成的信息互补感知注意力模块 (ICPAM). 全局注意力和局部注意力并行运作, 相互补充, 共同提升定位精度. 门控特征调节器则通过精细嵌入从全局和局部特征图中精确过滤掉不相关的特征和噪声响应, 为下游识别和学习由不同篡改技术引发的异常纹理、边缘变化及其他特征信息提供有力支持.

(3) 最后, 本文设计了一种新的联合损失函数以进一步提高网络的检测和定位性能. 实验结果表明, 本文方法的检测准确率提高了 2.3%. 此外, 在鲁棒性和泛化性上同样表现出较好的性能.

1 相关工作

在数字化时代, 图像已成为互联网中传递信息的核心媒介, 其真实性与完整性对于新闻报道、科学研究、法律取证及众多其他领域均至关重要. 因此, 图像篡改的检测与定位技术一直是研究领域的热门话题. 当前, 图像篡改检测方法可以分为基于传统手段的检测方法和基于深度学习的检测方法[1].

在研究初期, 传统的图像篡改检测技术占据主导地位. 然而, 这些传统方法大多仅针对特定的篡改手段, 如复制粘贴篡改和拼接篡改[2]. 虽然传统检测方法针对特定篡改类型的图像表现出一定的有效性, 但篡改方法的多样性和复杂性却削弱了传统方法的适应性和普遍性. 在实际应用中, 这些传统方法遭遇了诸多挑战. 随着人工智能技术的不断演进, 深度学习方法的优势日益凸显, 为图像篡改检测带来了新的曙光.

基于深度学习的图像篡改检测方法能够识别包括复制粘贴、拼接、删除等在内的多种篡改手段, 显著增强了检测的通用性和鲁棒性. 其中, 一些方法借助注意力机制来优化检测效果, 例如, Hu 等人[3]提出了一种空间金字塔注意网络 (SPAN), 该网络通过局部自注意金字塔来细化多尺度图像模块之间的关系建模, 并整合了一个位置投影机制以编码这些模块的空间位置信息. 此外, 还有部分方法则专注于通过精确定位局部异常来检测篡改痕迹. 例如, Wu 等人[4]将局部异常检测作为图像篡改检测和定位的一大挑战, 开发了一种 Z-score 特征来捕获局部异常, 并提出了一种创新的 LSTM 来评估异常区域. 同样, Liao 等人[5]提出了一种双流序列取证卷积网络来捕获篡改伪影和局部噪声残留证据. 然而, 由于他们忽视了篡改边缘信息的关键作用, 导致最终的预测掩码效果相当粗糙. 鉴于仅依赖局部异常检测的局限性, 一些方法开始采用多尺度技术在更广泛的范围内检测异常, 例如, PSCC-Net[6]中设计了两个不同的路径: 自上而下的路径, 利用密集交叉网络特征提取, 和自下而上的路径, 采用非本地注意机制, 以从粗到细的方式在多个尺度上预测篡改区域, 并自适应地学习不同图像区域之间的相关性. 然而, 由于过分强调无关的特征, 产生的预测掩码有时会表现出噪声和模糊现象[7]. 接着, Li 等人[8]认为多数方法在伪造区域与真实区域间普遍存在严重的特征耦合问题, 为此提出了一个分为两步的边缘感知的区域消息传递控制策略 (ERMPC). 在第 1 步中, 通过构建上下文增强图来整合全局语义信息, 以此区分伪造区域与真实区域之间的特征差异, 并利用阈值自适应可二值化边缘算法, 在初步输出结果上生成可学习的边缘信息. 第 2 步, 在可学习边缘信息的引导下, 设计了一个区域信息传递控制器, 旨在削弱伪造区域与真实区域之间的信息交换. 此方法通过明确建模两者间的不一致性, 在篡改图像检测中展现出了优异的性能. 此外, 受信噪分离思想的启发, Chen 等人[9]将图像伪造检测的后处理问题重新定义为信噪分离问题, 并据此提出了一种基于盲信号分离网络 (SNIS) 的方法. 该方法不仅能够有效消除复杂背景及后处理操作对伪造定位的视觉干扰, 还通过并行的扩张卷积结构学习多尺度信息, 有效地将后处理操作与篡改特征信息分离, 从而提升了检测鲁棒性.

2 基于多尺度感知学习的图像篡改检测与定位网络 (MsPL-Net)

本文提出的 MsPL-Net 是一种通用型的图像篡改检测和定位方法, 且特别注重解决定位问题. 相较于篡改检测, 图像篡改区域的像素级定位更为复杂. 鉴于检测与定位特征的学习是相辅相成的, 因此提升定位性能的同时, 也能相应地增强检测效果. 如图 1 所示, 网络的体系结构由 3 个核心部分组成: 特征提取模块、特征融合模块以及检测定位模块.

本文受 PSCC-Net[6]的启发, 借鉴了其自上而下的渐进式特征学习与自下而上的特征融合理念. 在自上而下的路径中, 主干编码器在一个跨不同尺度的分层密集连接框架中使用多尺度扩展卷积模块 (MSDCM). 这种方法增强了信息的传输, 并有效地应对了尺度变化带来的挑战. 本文选用了轻量级的 HRNet[10]作为主干网络, 并确定了 4 尺度网络为最优配置, 其中每层尺度的比例因子为上一层的 1/2. 而在自底向上的路径中, 逐步引入了 ICPAM 模块, 以从小到大的方式收集局部与全局信息, 生成预测掩码. 每层的预测掩码以前一个尺度的预测掩码作为先验, 通过迭代的方式逐步细化预测, 直至达到最终尺度, 从而生成一个逐步精细化的最终掩码. 最后, 将学习到的特征输入专为图像篡改检测设计的检测头[6], 根据预测的分数进行二值分类, 以确定图像的真实性.

2.1 多尺度扩张卷积模块 (MSDCM)

“多尺度”是指在不同尺度上对数据进行采样, 以

捕获不同尺度下的特征信息. 一般在更大的尺度上, 图像经常被缩小规模以包含更广阔的背景. 然而, 图像被缩小的同时会伴随着空间分辨率的下降, 标准的卷积和连续的池化操作可能会影响细粒度的细节. 通常, 较高的分辨率意味着对可见细节更丰富的描述和更多的信息内容, 而较低的分辨率会导致细节的减少和信息的减少. 保持空间分辨率对于保持输入图像的像素级空间信息至关重要, 这直接影响到图像的细节丰富度和信息量. 这对于确保网络在处理多尺度信息时既不损失细节也不引入模糊信息具有决定性作用. 因此, 本文设计了 MSDCM 采用扩张卷积[11]进行特征提取, 同时保持输入图像的空间分辨率, 有效地拓宽了每层的接收域, 获取了更广泛的上下文信息. 此外, 扩展卷积通过引入间隙来处理图像数据, 使特征提取过程更加平滑且连续. 这种平滑且连续的提取方法增强了模型在处理图像中的边缘和复杂细节方面的能力, 从而最大限度地减少了错误检测的概率. 本文为扩张卷积设置不同的扩张率, 并内嵌到相应的尺度层中, 以学习不同尺度上的特征, 每层扩张卷积的扩张率设为上一尺度的 2 倍, 这有助于利用扩张卷积的优势在捕获图像多尺度信息的同时保持空间分辨率, 扩展接收域, 增强鲁棒性.

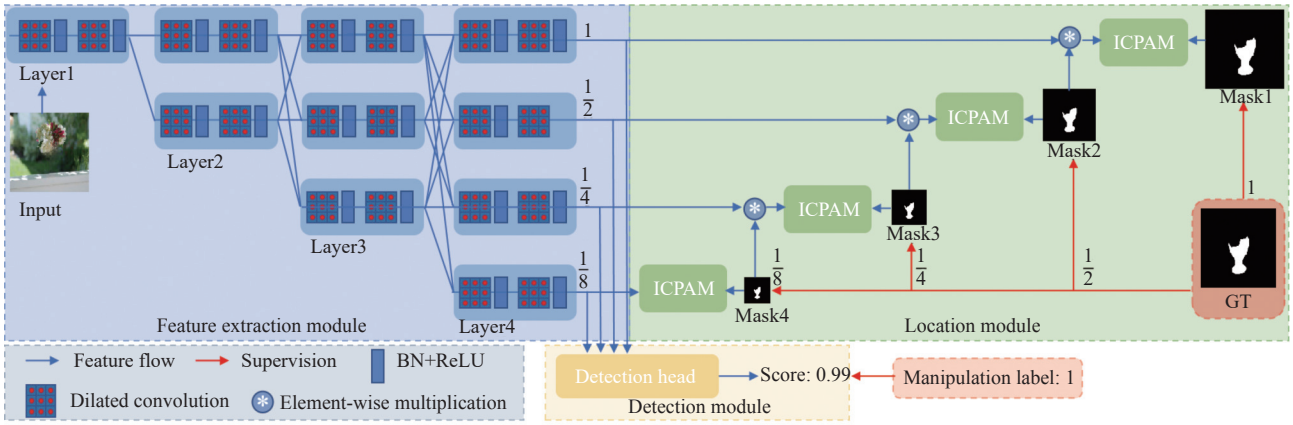

图 1 MsPL-Net 网络架构图

2.2 信息互补感知注意模块 (ICPAM)

在 ICPAM 中, 全局注意[6]将空间和通道注意机制相结合, 以捕捉特征之间复杂的空间和通道相关性. 空间注意机制基于上下文相关性聚集特征, 而通道注意机制则利用通道相关性进行特征聚合. 这种集成的方法增强了网络理解跨越各种范围的长距离依赖关系的能力[6]. 此外, 本文还设计了一个局部注意分支, 利用空间和通道注意机制来学习像素级的特征表示, 从而捕获局部显著的特征信息. 全局和局部两个注意模块并行运行, 其功能相互补充. ICPAM 利用全局关注为特征地图提供了总体线索, 而局部关注则强化了局部显著的特征, 最终提高了定位精度. ICPAM 的详细结构如图 2 所示.

因为即使是很小的特征图也可能会有极大地空间相关性, 很可能会超过内存限制[6], 所以先将输入的三维张量特征图 X 重塑为二维矩阵 X_1'. 这不仅保留了所有特征信息, 还避免了对潜在极端尺寸的特征图进行建模. 在构建全局空间和信道相关性之前, 我们使用 1×1 卷积来构建嵌入函数 F_1、F_2 和 F_3, 并分别生成嵌入式特征 $X_a = F_1(X_1')$、$X_b = F_2(X_1')$ 和 $X_c = F_3(X_1')$. 随后, 利用嵌入的特征 X_b 和 X_c 计算空间注意系数 K_{1s} 和通道注意系数 K_{1c}, 计算公式如下:

$$K_{1s} = Softmax(X_{1\theta}' X_{1\varphi}'^{\mathrm{T}}),\ K_{1s} \in R^{(H_q \times W_q)/r^2 \times (H_q \times W_q)/r^2} \quad (1)$$

$$K_{1c} = Softmax(X_{1\theta}'^{\mathrm{T}} X_{1\varphi}'),\ K_{1c} \in R^{Cr^2 \times Cr^2} \quad (2)$$

在获得空间和通道注意系数后, 通过矩阵乘法实现空间和通道注意机制, 两者都通过相同的线性嵌入 $X_a = F_1(X_1')$ 以进行相互调节. 这些机制基于注意力系数聚合特征图, 依据注意力系数增强伪造区域中的表示, 系数值越高即相关性越高. 这有助于网络学习区分伪造区域和原始区域的特征表示. 值越高表明相关性越大. 全局空间注意特征图 Y_{1s}' 和全局通道注意特征图 Y_{1c}' 计算公式如下:

$$Y'_{1s}=K_{1s}X_{\alpha},\ Y'_{1s}\in R^{(H_q\times W_q)/r^2\times Cr^2} \tag{3}$$

$$Y'_{1c}=X_{\alpha}K_{1c},\ Y'_{1c}\in R^{(H_q\times W_q)/r^2\times Cr^2} \tag{4}$$

其中, K_{1s}中的元素 (i,j) 表示$X'_{1\theta}$的第 i 行和$X'_{1\varphi}$的第 j 行中的特征向量之间的相似性. K_{1c}中的元素 (i,j) 表示$X'_{1\theta}$的第 i 列和$X'_{1\varphi}$的第 j 列中的通道图之间的相似性. ρ_s和ρ_c随后, 将Y'_{1s}、Y'_{1c}重塑为原来尺寸Y_{1s}、Y_{1c}, 并集成到全局注意力特征图Z_1中, 计算公式如下:

$$Z_1=\alpha_s\times\rho_s\times Y_{1s}+\alpha_c\times\rho_c\times Y_{1c},\ Z_1\in R^{H\times W\times C} \tag{5}$$

其中, ρ_s和ρ_c是由 1×1 卷积构造的两个函数, 它们的输出是互补的, 以改进特征表示. α_s和α_c ($\geqslant$0) 是两个可训练的参数, 它们自适应地权衡了空间和通道注意力对输入数据中更相关更具代表性部分的关注.

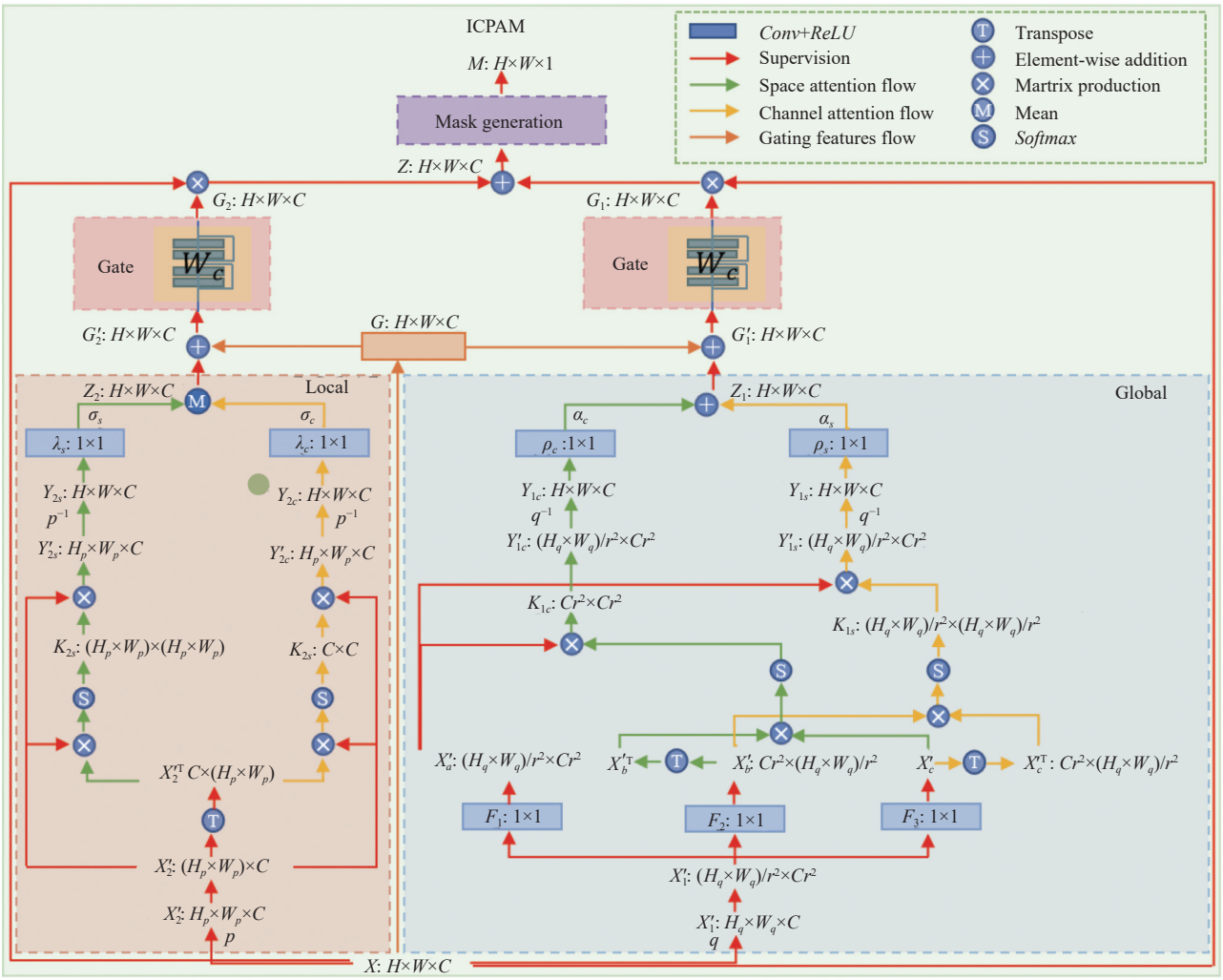

图 2 ICPAM 模块详细设计图

局部注意分支同样首先将输入张量通过池化因子 p 输入到最大池层, 得到的降采样张量, 将三维张量 X 转换为二维矩阵X'_2. 为了提高注意的准确性, 局部注意分支从X'_2中提取相关特征, 并对空间和通道相关性进行建模. 空间注意系数K_{2s}和通道注意系数K_{2c}计算公式如下:

$$K_{2s}=Softmax\left(\frac{1}{\sqrt{H_p\times W_p}}X'_2X'^{\mathrm{T}}_2\right),\ K_{2s}\in R^{(H_p\times W_p)\times C} \tag{6}$$

$$K_{2c}=Softmax\left(\frac{1}{C}X'^{\mathrm{T}}_2X'_2\right),\ K_{2c}\in R^{C\times C} \tag{7}$$

其中, T 表示转置运算, $1/\sqrt{H_p\times W_p}$和 $1/C$ 为归一化因子用来限制乘积输出的大小, 从而确保 *Softmax* 函数的

稳定性, C 是 X_2' 的通道数量. 随后, 通过矩阵乘法实现注意力关注, 空间注意特征图 Y_{2s}' 和通道注意特征图 Y_{2c}' 计算公式如下:

$$Y_{2s}' = K_{2s}X_2',\ Y_{2s}' \in R^{(H_p \times W_p)\times C} \tag{8}$$

$$Y_{2c}' = K_{2c}X_2',\ Y_{2c}' \in R^{(H_p \times W_p)\times C} \tag{9}$$

再将输入的 Y_{2s}'、Y_{2c}' 重塑为原来大小 Y_{2s}、Y_{2c}. 与全局注意力特征图 Z_1 类似, 局部注意力特征图 Z_2 计算公式如下:

$$Z_2 = \sigma_s \times \mu_s \times Y_{2s} + \sigma_c \times \mu_c \times Y_{2c},\ Z_2 \in R^{H\times W\times C} \tag{10}$$

为了将更精确的更具鉴别性的特征传播到下游, 本文通过门控特征调节器 (Gate) 使用精细嵌入函数来过滤不相关特征和噪声响应, 使得下游进一步识别学习不同篡改手段引起的异常纹理, 边缘变化及异常噪声分布等特征.

为了得到门控系数, 首先将特征张量 Z_i 和门控特征张量 G 转换为中间向量 G_i'. 门控特征 G 是由输入特征 X 上采样得到的特征. 中间向量 G_i' 计算公式如下:

$$G_i' = (Z_i + G),\ i = 1,2,\ G_i' \in R^{H\times W\times C} \tag{11}$$

然后, Gate 使用非线性变换层对该向量进行重新采样生成各注意力的门控特征 G_i, 计算公式如下:

$$G_i = ReLU(Conv(G_i')),\ i = 1,2,\ G_i \in R^{H\times W\times C} \tag{12}$$

最后, 采用残差学习结合 G_1、G_2 和输入特征 X 表示最终特征图 Z, 为了减少特征冗余和减少对不相关信息的过度强调, 我们为局部注意特征图分配了 0.5 的权重. 特征图 Z 计算公式如下:

$$Z = X + G_1 + \frac{1}{2} \times G_2, Z \in R^{H\times W\times C} \tag{13}$$

在 ICPAM 最终采用掩码生成块减少 Z 的通道数, 输出是一个一通道的预测掩码 M, 计算公式如下:

$$M = Conv(ReLU(Conv(Sigmoid(Z)))),\ M \in R^{H\times W\times 1} \tag{14}$$

2.3 损失函数

本文提出了一种结合检测与定位损失的多任务联合损失函数, 该函数包括二元交叉熵损失函数 (BCE loss)、Dice 系数差异函数 (Dice loss) 和焦点损失 (Focal loss).

BCE 损失计算预测值与真实值之间的交叉熵, 通过给错误的预测施加更大的惩罚来指导模型调整其参数, 使得预测概率更接近真实标签. 它鼓励模型在每个像素位置上做出准确的分类决策, 从而提高分类精度. BCE loss 计算公式如下:

$$L_{\text{bce}} = -\frac{1}{N}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \tag{15}$$

其中, $\hat{y}_i$ 是预测值, $\hat{y}_i \in [0,1]$; y_i 是实际标签.

Dice 损失基于 Dice 系数来衡量预测结果与真实结果的相似程度. Dice 系数是预测区域和真实区域的交集与它们的并集之比. 在图像篡改定位中, 我们可以将定位篡改区域视为一种特殊的图像分割任务, 模型通过最小化 Dice 损失, 可以获得更好的逐像素分割性能, 更精确地分割出篡改区域, 提高定位的准确性. Dice loss 计算公式如下:

$$L_{\text{dice}} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{16}$$

如果算法过度依赖于训练中某一特定类型的篡改数据, 而未能广泛学习到其他类型的篡改数据, 就会导致过拟合, 从而使模型不能很好地泛化到未见过的数据. 焦点损失通过降低易分类样本的权重来关注难分类样本, 从而平衡不同篡改类型的处理能力, 使模型在面对不同类型的篡改时都能保持较好的性能. Focal loss 计算公式如下:

$$L_{\text{focal}} = -\alpha_t(1-P_t)^{\gamma}\log(P_t) \tag{17}$$

BCE 损失用于指导图像分类, Dice 损失用于指导图像定位, 焦点损失用于平衡不同篡改类型的处理能力, 提高泛化性. 由于最终的定位预测 Mask 由 4 个不同的尺度预测的 Mask: G_4、G_3、G_2 和 G_1 由粗到细逐渐精确确定, 并以完全监督的形式对每个尺度进行明确监督, 我们视 4 个尺度同等重要. 在这些 Mask 中, 0 代表真实像素, 1 代表伪造像素. 我们最终的损失函数计算公式如下:

$$L = L_{\text{bce}}(s_d, l_d) + \frac{1}{4}\sum_{m=1}^{4} L_{\text{dice}}(M_m, G_m) + L_{\text{focal}} \tag{18}$$

3 实验分析

3.1 实验设置

(1) 预训练数据集: 包括拼接组合, 复制粘贴, 移动删除和真实 4 种类型的图像. 每种图像类型的生成方法如下: 使用 MS COCO[12]数据集补充图像训练集. 此外, 从 Viseon[13]和 Dresden[14]等数据集中随机选择供体

和目标图像, 并采用类似的操作来补充图像训练集.

(2) 测试数据集: 使用 Coverage[15]、Columbia[16]、CASIA[17]、NIST2016[18]和 IMD2020[19] 这 5 个公开的数据集来评估本方法的有效性. 表 1 总结了每个数据集的详细信息, 以及用于评估预训练模型和微调模型的图像数量.

表 1 使用数据集介绍表

数据集	预训练模型	微调模型		复制粘贴	拼接组合	移动删除	图像格式
	#训练	#测试	#训练				
Columbia	180	450	114	√	×	×	PNG
Coverage	100	80	20	×	√	×	PNG
CASIA	6 044	5 123	921	√	√	×	TIFF, JPEG
NIST2016	564	—	—	√	√	√	PNG
IMD2020	2 010	—	—	√	√	√	PNG

(3) 评估指标: 本文在定位性能评估方面, 采用了 *AUC*、*F*1-score (*F*1)、*IoU* 以及 *MCC* 等指标, 而在检测性能评估方面, 则选用了 *AUC*、*F*1-score、等错误率 (*EER*) 和 1% 的假阳性率 (*TRP*) 指标进行性能评估. *AUC* 是 *ROC* 曲线下的面积, 主要用于评估模型在分类任务中的性能. *F*1-score 通过结合精度和召回率提供了一个平衡的评估. *IoU* 衡量的是两个区域 (通常是预测框和真实框) 之间的重叠程度, 因此它可以用来度量预测的篡改区域和实际篡改区域之间的重叠程度, 从而评估模型在定位篡改区域时的准确性. 而 *MCC* 则考虑了真正例、假正例、假负例、真负例这 4 个类别的预测结果, 因此它能够衡量预测框的全面性. 等误率 (*EER*) 是错误接受率 (*FAR*) 等于错误拒绝率 (*FRR*) 的错误率. 较低的 *EER* 表示模型在区分篡改和未篡改图像方面的性能更好. 最后, 真阳性率 (*TPR*) 测量了模型准确识别的实际篡改实例的比例, 反映了其在保持低假阳性率的同时精确检测篡改区域的能力. *F*1、*TRP*、*IoU* 与 *MCC* 计算公式如下:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{19}$$

$$TRP = \frac{TP_{(FPR=1\%)}}{TP_{(FPR=1\%)} + FN_{(FPR=1\%)}} \tag{20}$$

$$IoU = \frac{\text{交集区域的面积}}{\text{并集区域的面积}} \tag{21}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{22}$$

(4) 实施细节: 本文模型采用 PyTorch 实现, 使用 NVIDIA Tesla P40 24 GB 显卡进行训练. 因为本文选用的 HRNet[10]主干的低层卷积结构与 ImageNet 兼容, 可以直接加载其预训练权重初始化本文网络主干, 并通过 Adam[20]优化整个模型, 目的为任务提供良好的起点, 以便提升训练效果和效率. 批量大小为 10, 初始学习率为 2E−4. 学习率每 5 个周期减半, 总训练周期为 30 个周期.

3.2 检测性能比较

由于 ManTra[4]和 SPAN[3]方法没有直接进行检测评估, 我们使用他们预测的篡改掩码的平均值作为他们的分数. 表 2 给出了各种方法在 CASIA-D 数据集上的检测评价比较. 我们的方法在所有采用的指标中都取得了最好的检测性能.

表 2 检测性能评估表 (%)

方法	*AUC*↑	*F*1↑	*EER*↓	*TPR*↑
ManTra[4] (avg)	59.94	56.69	43.21	5.43
SPAN[3] (avg)	67.33	63.48	36.47	5.54
PSCC-Net[6] (avg)	74.40	66.88	33.21	28.37
PSCC-Net[6]	99.65	97.12	2.83	95.65
Our (avg)	**81.07**	**72.23**	**21.39**	**37.86**
Our	**99.89**	**98.63**	**1.93**	**98.23**

3.3 定位性能比较

根据 SPAN[3]中定义的评估协议, 本文采用预训练模型和微调模型来比较图像篡改检测与定位的性能. 预训练模型可以体现模型的泛化能力, 微调模型则体现定位性能. 由于一些比较网络没有微调的模型, 我们使用预训练模型进行检测评估比较.

(1) 预训练模型比较: 表 3 中比较了不同方法的预训练模型在像素级 *AUC* 下的定位性能. 从表 3 中可知, 本文的预训练模型在 Columbia、NIST2016、CASIA 和 IMD2020 数据集上均取得了最佳的定位性能. 尤其在 IMD2020 上性能提升最高, 这表明 MsPL-Net 相较于其他方法具有更好的泛化能力.

表 3 预训练模型定位性能 *AUC* 评估 (%)

数据集	ManTra[4]	SPAN[3]	PSCC-Net[6]	ObjectFormer[21]	ERMPC[8]	Our
Columbia	82.4	93.6	98.2	95.5	96.8	**99.1**
Coverage	81.9	92.2	84.7	92.8	**94.4**	91.7
CASIA	81.7	79.7	82.9	84.3	87.6	**87.8**
NIST2016	79.5	84.0	85.5	87.2	89.5	**89.9**
IMD2020	74.8	75.0	80.6	82.1	85.6	**87.5**

(2) 微调模型比较: 微调模型的训练策略与预训练模型相同, 预训练模型的网络权重被用于初始化微调模型. 在 Coverage、CASIA 和 NIST2016 数据集上的训练分割进行微调并调整设置初始学习率为 1E−4. 表 4 展示了 CASIA 数据集上 *IoU* 和 *MCC* 两个指标的定位性能比较, 可以看到所提方法的性能都是最优的, 这说明了本文方法预测的篡改区域不仅与实际篡改区域有着良好的重叠度, 而且还有着较好的全面性. 表 5 展示了 CASIA 和 NIST2016 数据集上 MsPL-Net 在 *AUC* 和 *F*1-score 指标上都依旧是最优, 这也验证了我们整体网络设计的优越性.

表 4 微调模型在 *IoU* 与 *MCC* 指标下的性能评估

指标	ManTra[4]	DenFCN[22]	IECG[23]	PSCC-Net[6]	Our
IoU	0.36	0.55	0.44	0.58	**0.62**
MCC	0.45	0.62	0.49	0.65	**0.68**

(3) 定位可视化比较: 在图 3 中, 展示了针对不同篡改类型和不同篡改尺度图片的可视化定位结果. 相较于其他方法的网络, 所提网络的预测掩码具有更高的精度, 更少的像素误报, 边缘也更清晰. 图 4 则展示了对 CASIA-D 数据集定位篡改区域的可视化结果, 将 PSCC-Net[6]的预测掩码与本文方法对原始真实图像和篡改图像的 GT 进行了比较. 对于原始的未篡改图像, 本文网络生成的预测掩码几乎完美无瑕, 而 PSCC-Net[6] 则产生大量的误报. 至于篡改图像, MsPL-Net 同样展现出了超越其他方法的卓越性能.

表 5 微调模型在 *AUC* 与 *F*1 指标下的性能评估 (%)

方法	Coverage	CASIA	NIST2016
J-LSTM[24]	61.4/—	—/—	76.4/—
H-LSTM[25]	71.2/—	—/—	79.4/—
RGB-N[26]	81.7/43.7	79.5/40.8	93.7/72.2
SPAN[3]	93.7/55.8	83.8/38.2	96.1/58.2
PSCC-Net[6]	94.1/72.3	87.5/55.4	99.1/74.2
ERMPC[8]	**98.4/77.3**	90.4/58.6	99.7/83.6
Our	98.2/76.7	**91.7/60.1**	**99.8/85.2**

注: /前后分别为*AUC*和*F*1值.

(a) 复制粘贴 (b) 移动删除 (c) 拼接组合

图 3 定位性能可视化图

3.4 尺度选择和消融实验

为了确定本文网络模型峰值性能的尺度层的最佳数量, 我们在 Columbia、Coverage 和 NIST2016 数据集上, 针对 3 种不同的尺度层配置进行了严格的实验评估. 如图 5 所示, 实验结果表明, 当网络配置为 4 个尺度层时, 其检测性能达到了最优.

网络模块消融实验中, 我们对网络的几个变种进行了测试, 以验证网络设计的合理性. 通过表 6 网络模块消融实验可以看出, 我们提出的 MSDCM、ICPAM 和 Loss 对于提升网络模型性能都是有益的.

此外, 本文还针对损失函数进行消融实验, 并利用收敛曲线来评估损失函数设计的合理性. 如表 7 所示, 通过结合二元交叉熵损失 (BCE loss)、Dice 系数差异函数 (Dice loss)、焦点损失 (Focal loss), 本文的损失函数使得模型达到了最优状态. 如图 6 所示的训练损失和验证损失的收敛曲线, 为我们提供了损失函数在训练过程中的直观表现. 从图 6 中可以看出, 随着训练轮次的增加, 损失逐渐减小并趋于稳定, 说明

我们的损失函数在训练过程中有效地引导模型优化. 尤其是在 25 轮次之后, 模型的损失值已经趋近于最小值, 表明损失函数在多轮迭代中有效收敛, 达到了最优性能.

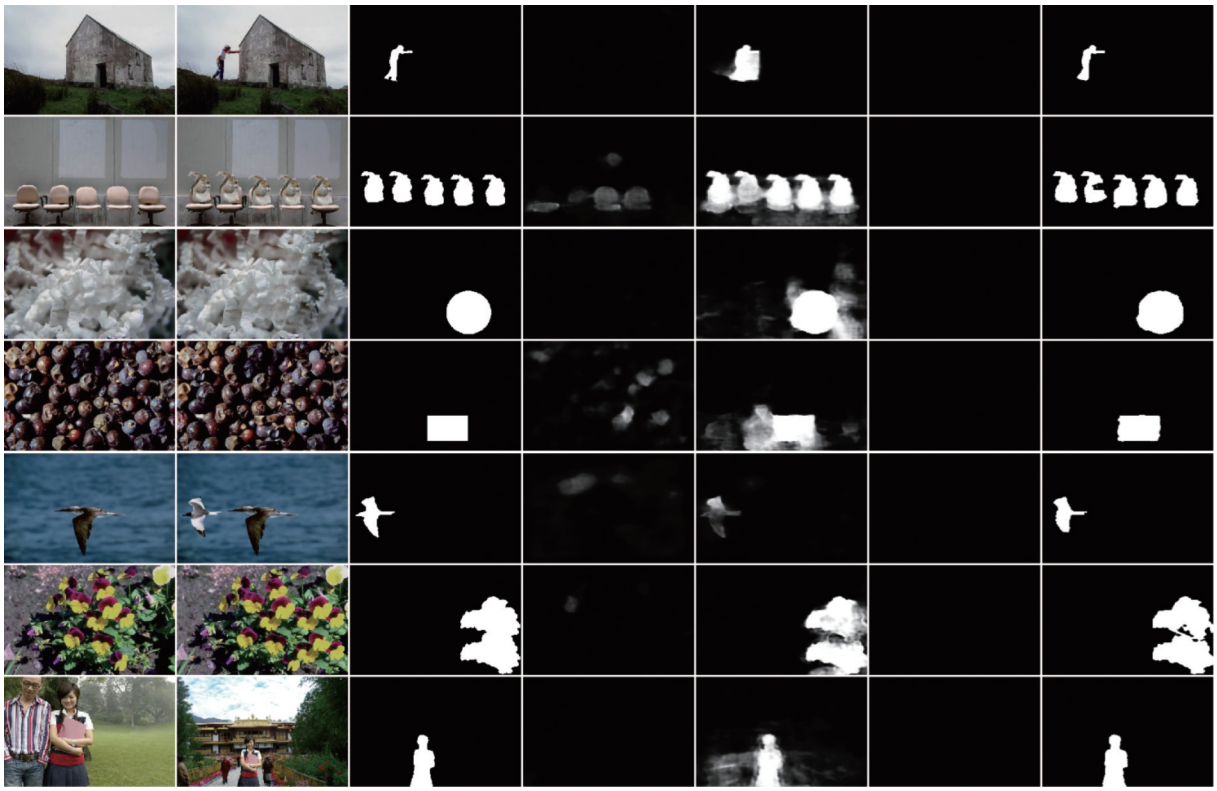

图 4 CASIA-D 数据集定位性能比较可视化图

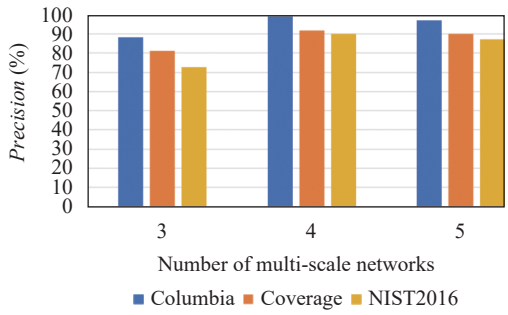

图 5 尺度选择实验图

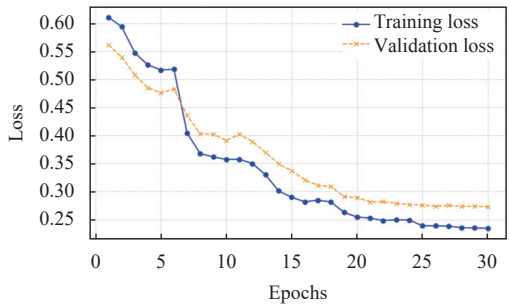

图 6 损失函数收敛曲线图

表 6 网络模块消融实验 (*AUC*) (%)

MSDCM	ICPAM	Loss	Coverage	CASIA
√	×	×	67.9	63.2
√	√	×	78.7	74..6
√	√	√	**94.4**	**87.8**

表 7 损失函数消融实验 (*AUC*) (%)

BCE loss	Dice loss	Focal loss	NIST2016	Columbia
×	×	×	72.9	78.8
√	×	×	79.4	83.7
√	√	×	87.5	98.6
√	√	√	**89.9**	**99.1**

3.5 不同尺度上预测操纵 Mask 的可视化

图 7 展示了 ICPAM 从尺度 4 到尺度 1 的预测掩码的逐步可视化结果. 具体地说, Mask4、Mask3 和 Mask2 分别表示在尺度 4、3 和 2 上预测的中间掩码, 最终得到最终 Mask1. 可以观察到, 从 Mask4 到 Mask1, 篡改区域的定位逐渐变得更加清晰且完整, 呈现出显著的优化趋势. 这一现象进一步证实了我们对于网络尺度层数的精心选择以及 ICPAM 设计的有效性和合理性.

图 7 Mask4 至 Mask1 的预测掩码可视化图

3.6 鲁棒性分析

为了研究网络定位的鲁棒性, 我们遵循 SPAN[3]与 PSCC-Net[6]中的后处理设置, 对 NIST2016 的原始篡改图像进行后处理. 这些操作包括通过因子 S 对图像进行缩放 (Resize), 引入核大小为 K 的高斯模糊 (GSBlur)、添加标准差 P 的高斯噪声 (GSNoise), 以及使用质量因子 Q 进行 JPEG 压缩 (JPEGComp). 表 8 展示了预训练模型在各种失真情况下像素级 AUC 的定位鲁棒性分析情况. 在后处理干扰的情况下, 我们的网络表现出优于其他网络的鲁棒性.

表 8 鲁棒性评估 (%)

后处理操作	Resize		GSBlur		GSNoise		JPEGComp		Mixed	w/o distortion
	S=0.78	S=0.25	K=3	K=15	P=3	P=15	Q=100	Q=50		
ManTra[4]	77.43	75.52	77.46	74.55	67.41	58.55	77.91	74.38	64.82	78.05
SPAN[3]	83.24	80.32	83.10	79.15	75.17	67.28	83.59	80.68	68.36	83.95
PSCC-Net[6]	85.29	85.01	85.38	79.93	78.42	76.65	85.40	85.37	73.93	85.47
ObjectFormer[21]	87.17	86.33	85.97	80.26	79.58	78.15	86.37	86.24	—	87.18
ERMPC[8]	89.33	87.72	89.22	87.13	88.25	**83.40**	89.42	88.82	—	—
Our	**89.97**	**88.97**	**89.76**	**87.64**	**88.71**	82.98	**89.33**	**89.12**	**88.43**	**89.17**

4 结论与展望

本文设计了一个多尺度感知学习的图像篡改检测与定位网络 (MsPL-Net). 其具体来说, 我们设计了一个多尺度扩展卷积模块 (MSDCM), 以解决由不同的后处理和操作类型导致的弱特征鲁棒性问题. 此外, 还加入了一个由全局注意、局部注意组成的信息互补感知注意模块 (ICPAM) 和一个门控特征调节器, 以解决由于对篡改大小的敏感性而导致的篡改边缘位置模糊的问题. 该方法旨在有效地解决定位篡改区域边缘模糊的问题并提高鲁棒性, 为图像篡改检测与定位领域的进一步研究提供有力的支持.

参考文献

1 褚莹娜. 基于深度学习方法的图像篡改检测与定位研究 [硕士学位论文]. 天津: 天津理工大学, 2024.
2 宋炎寒, 苏红旗. 图像篡改检测技术综述. 电子技术与软件工程, 2021(21): 103–105.

3 Hu XF, Zhang ZH, Jiang ZY, *et al*. SPAN: Spatial pyramid attention network for image manipulation localization. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 312–328.
4 Wu Y, AbdAlmageed W, Natarajan P. ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 9535–9544.
5 Liao X, Li KD, Zhu XS, *et al*. Robust detection of image operator chain with two-stream convolutional neural network. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(5): 955–968. [doi: 10.1109/JSTSP.2020.3002391]
6 Liu XH, Liu YJ, Chen J, *et al*. PSCC-Net: Progressive spatio-channel correlation network for image manipulation detection and localization. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(11): 7505–7517. [doi: 10.1109/TCSVT.2022.3189545]
7 Li FY, Zhai HJ, Zhang XP, *et al*. Image manipulation localization using spatial-channel fusion excitation and fine-grained feature enhancement. IEEE Transactions on Instrumentation and Measurement, 2023, 73: 3500714.
8 Li D, Zhu JY, Wang ML, *et al*. Edge-aware regional message passing controller for image forgery localization. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 8222–8232.
9 Chen JX, Liao X, Wang W, *et al*. SNIS: A signal noise separation-based network for post-processed image forgery detection. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(2): 935–951. [doi: 10.1109/TCSVT.2022.3204753]
10 Wang JD, Sun K, Cheng TH, *et al*. Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3349–3364. [doi: 10.1109/TPAMI.2020.2983686]
11 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. Proceedings of the 4th International Conference on Learning Representations (ICLR). San Juan: OpenReview.net, 2016.
12 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 740–755.
13 Shullani D, Fontani M, Iuliani M, *et al*. VISION: A video and image dataset for source identification. EURASIP Journal on Information Security, 2017, 2017(1): 15. [doi: 10.1186/s13635-017-0067-2]
14 Gloe T, Böhme R. The 'Dresden Image Database' for benchmarking digital image forensics. Proceedings of the 2010 ACM Symposium on Applied Computing. Sierre: Association for Computing Machinery, 2010. 1584–1590.
15 Wen BH, Zhu Y, Subramanian R, *et al*. COVERAGE—A novel database for copy-move forgery detection. Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP). Phoenix: IEEE, 2016. 161–165.
16 Ng TT, Chang SF. A data set of authentic and spliced image blocks [Technical Report], #203-2004-3. New York: Columbia University, 2004. https://www.ee.columbia.edu/ln/dvmm/publications/04/TR_splicingDataSet_ttng.pdf
17 Dong J, Wang W, Tan TN. CASIA image tampering detection evaluation database. Proceedings of the 2013 IEEE China Summit and International Conference on Signal and Information Processing. Beijing: IEEE, 2013. 422–426.
18 Guan HY, Kozak M, Robertson E, *et al*. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation. Proceedings of the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). Waikoloa: IEEE, 2019. 63–72.
19 Novozámský A, Mahdian B, Saic S. IMD2020: A large-scale annotated dataset tailored for detecting manipulated images. Proceedings of the 2020 IEEE Winter Applications of Computer Vision Workshops. Snowmass: IEEE, 2020. 71–80.
20 Kingma DP, Ba J. Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR). San Diego: OpenReview.net, 2015.
21 Wang JK, Wu ZX, Chen JJ, *et al*. ObjectFormer for image manipulation detection and localization. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 2354–2363.
22 Zhuang PY, Li HD, Tan SQ, *et al*. Image tampering localization using a dense fully convolutional network. IEEE Transactions on Information Forensics and Security, 2021, 16: 2986–2999. [doi: 10.1109/TIFS.2021.3070444]
23 Lu W, Xu WB, Sheng ZQ. An interpretable image tampering detection approach based on cooperative game. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(2): 952–962. [doi: 10.1109/TCSVT.2022.3204740]
24 Bappy JH, Roy-Chowdhury AK, Bunk J, *et al*. Exploiting spatial structure for localizing manipulated image regions. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 4980–4989.
25 Bappy JH, Simons C, Nataraj L, *et al*. Hybrid LSTM and encoder-decoder architecture for detection of image forgeries. IEEE Transactions on Image Processing, 2019, 28(7): 3286–3300. [doi: 10.1109/TIP.2019.2895466]
26 Zhou P, Han XT, Morariu VI, *et al*. Learning rich features for image manipulation detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1053–1061.

(校对责编: 张重毅)