

文章编号:1007-2780(2024)08-1103-13

基于双流增强编码和注意优化解码的 图像篡改定位算法

朱 叶, 赵晓祥, 于 洋*

(河北工业大学 人工智能与数据科学学院, 天津 300401)

摘要:主流图像篡改定位方法通常通过简单操作融合不同流的不一致特征,导致特征冗余且篡改区域的像素误检。基于此,本文提出基于双流增强编码和注意优化解码的图像篡改定位框架。首先,提出双流增强编码分别对图像和频域特征进行基于噪声和通道注意力的自增强和基于特征映射的交叉注意权重的交互增强。随后,引入多级感受野策略探索多尺度上下文信息,设计邻阶特征聚合模块融合多尺度相邻特征。最后,利用篡改区域和非篡改区域协同增强模型的篡改定位能力,提出注意优化解码模块,消除初始篡改区域预测中边缘像素的错误预测,逐步精确细化篡改定位。在4个主流公共基准数据集NIST16、Coverage、Columbia、CASIA和两个现实挑战数据集IMD20、Wild上与主流篡改定位方法进行对比,本文算法在无微调模型和微调模型两个设置下,在6个数据集上的性能最优,证明本文提出的篡改定位网络能够充分利用多种篡改线索,在不同的篡改数据集上实现篡改区域的有效定位,具有更高的定位精度和更强的鲁棒性。

关键词:图像篡改定位;双流增强编码;注意优化解码;邻阶特征聚合

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2023-0280

Dual-stream enhancement encoder and attention optimization decoder for image manipulation localization

ZHU Ye, ZHAO Xiaoxiang, YU Yang*

(School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China)

Abstract: Mainstream image manipulation localization methods usually fuse inconsistent features of different streams through simple operations, resulting in feature redundancy and pixel misdetection of tampered regions. Therefore, we propose a novel network of dual-stream enhancement encoder and attention optimization decoder for image manipulation localization. Firstly, the dual-stream enhancement encoder module can self-reinforce and interact with the extracted dual-stream multi-scale features, and can make full use of a variety of tampered information, so that a variety of tampered information can be complemented by interaction, and more attention is paid to the tampering features. Then, a multi-scale receptive field strategy is introduced

收稿日期:2023-08-28;**修订日期:**2023-10-02.

基金项目:国家自然科学基金青年项目(No.62102129);河北省自然科学基金(No.F2021202030);石家庄市驻冀高校基础研究项目(No.241790817A)

Supported by Youth Program of National Natural Science Foundation of China (No.62102129); Natural Science Foundation of Hebei Province (No.F2021202030); Basic Research Project of Hebei Universities in Shijiazhuang (No.241790817A)

*通信联系人, E-mail: yuyang@hebut.edu.cn

to explore multi-scale context information, and an adjacent-level feature aggregation module is designed to fuse multi-scale adjacent features. Finally, the capability of manipulation localization is enhanced with the cooperation of tamper region and genuine region, the attention optimization decoder module is designed to eliminate the wrong prediction of edge pixels in the initial tamper region prediction, and the manipulation localization is refined step by step. Extensive experiments are constructed on four mainstream public datasets, NIST16, Coverage, Columbia and CASIA, and two realistic challenge datasets, IMD20 and Wild, to compare with mainstream manipulation localization methods. Our proposed method has superior performance under six datasets in the settings of none fine-tuning and fine-tuning model, which demonstrates that our proposed method can make full use of various forgery clues to achieve greater localization accuracy and stronger robustness.

Key words: image manipulation localization; dual-stream enhancement encoder; attention optimization decoder; adjacent feature aggregation module

1 引 言

随着 Photoshop、美图秀秀等数字图像编辑工具的发展和普及,人们可以低成本、无痕迹地对数字图像进行美化和编辑。图像模糊、对比度增强和颜色调整等图像美化操作不修改图像内容。但是对图像内容进行更改的图像操作若被用于不道德和非法目的,如军事通信、政治新闻照片、法律证据、电子账单、企业资质等场景,会对信息安全和社会稳定造成巨大危害。一般来说,主流图像内容篡改操作分为以下 3 类:拼接^[1]、复制-粘贴^[2]和移除^[3]。拼接是将真实图像目标区域复制到另一幅目标图像上,复制-粘贴是将真实图像目标区域复制-粘贴到同幅图像中,移除技术先移除图像目标区域再进行修复。隐藏篡改痕迹,噪声、模糊和重压缩等后处理操作被应用于篡改图像,加大了篡改区域定位的难度。

近年来,图像篡改检测的问题变得越来越具有挑战性,并得到了相当多的关注。在图像篡改检测方面有很多方法被不断提出。早期篡改取证方法主要提取手工设计特征,如相机滤波阵列(Color Filter Array, CFA)模式^[4]、不一致的局部噪声^[5]、JPEG 量化表不一致^[6]等。该类特征在图像区域篡改后保持不变,从而可以区分图像真实和篡改区域,但大多仅能检测单一的篡改操作。随着大量复杂图像操作的出现,以上方法已不适合实际应用。

不同类型的篡改操作具有不同的特征。多类型篡改取证较单一,篡改更为复杂。近年来,

随着深度学习技术在图像视觉任务中的广泛应用,其在显著性目标检测任务^[7-8]中表现出了良好的性能。因此,研究者认为可以借鉴深度学习技术,解决图像篡改定位任务。深度神经网络的图像篡改检测方法都取得了成功,如基于深度网络的篡改定位^[9]、基于像素级的篡改定位^[10-11]和基于对象检测的篡改定位^[12]等。但是,神经网络更善于通过理解视觉上下文来区分图像内容,而图像显著内容可能与图像篡改定位无关,经过篡改的区域经常包含无意义的图像内容。

为了解决上述问题,Bappy 等人^[13]提出一种编码器-解码器结构来实现图像篡改检测。由于编码器和解码器之间只有信息传输,没有信息交换,不可避免地会造成特征信息的丢失。Chen 等人^[14]在编码器和解码器之间引入类似 U-NET 架构的跳跃层连接,可以保证特征不会随着解码器的传输消失,但仅在对尺度之间进行跳层,没有考虑其他尺度。主流图像篡改方法通常通过融合不同数据流的不一致特征训练定位模型,但其融合方式大多简单直接,造成深层特征冗余且篡改区域误检。基于此,本文提出基于双流增强编码和注意优化解码的图像篡改定位框架,主要贡献如下:

(1) 本文提出基于图像流和噪声流的双流增强编码模块,对图像和频域特征进行基于噪声和通道注意力的自增强和基于特征映射的交叉注意权重的交互增强,补充和增强不同篡改操作特征。

(2) 在解码阶段,本文提出篡改区域和非篡改区域协同增强模型的篡改定位能力,提出注意

优化解码模块,消除初始篡改区域预测中边缘像素的错误预测,逐步精确细化篡改定位。

(3)本文在6个公共数据集上与主流方法进行对比实验,实验结果表明,本文方法在篡改定位精度和模型鲁棒性方面都优于主流篡改定位方法。

2 相关工作

2.1 图像篡改定位

随着深度学习技术的发展和普及,人们试图通过引入深度神经网络构建一个通用的、端到端的、高效的图像处理定位框架。目前,基于深度学习的图像篡改定位模型主要使用噪声视图结合RGB特征或频域特征的方式来提高定位精度。Zhou等人^[9]提出基于RGB流和噪声流的双流Faster R-CNN网络以检测图像篡改区域。但该方法仅能获得粗略的篡改定位框,且对两流特征只是通过池化层进行交互。Li等人^[15]开发了一个由3个流组成的跨域架构,但只考虑了最后一个卷积层的输出特征,忽略了其他层特征的有效性,且只使用双线性池化层来融合三流特征。Wang等人^[12]提出ObjectFormer,利用RGB特征和频率特征的组合以及基于注意力的可学习对象原型来识别和定位篡改图像。但其对两流仅采用简单的连接操作来融合相同分辨率特征,没有实现两种篡改线索的充分交互补充。Lin等人^[16]提出了一种新的网络学习和增强多篡改痕迹,但只通过连接操作对噪声流和RGB流进行融合。

噪声视图利用拼接或移除引入的新元素与原始部分在噪声分布上不同的线索捕捉图像伪造痕迹,以检测篡改区域。噪声映射首先由预定义的高通滤波器或约束卷积层生成,然后将其单独^[17]或与RGB流一起^[18]馈送到深度神经网络。从这些噪声流中获得的噪声不一致性有助于增强图像的篡改痕迹。然而,如果不引入新元素,这种类型的方法在检测复制-粘贴时无效的。相比之下,频率信息主要通过离散余弦变换或快速傅里叶变换来提取,以方便捕捉在RGB域^[19]中不再可见的细微伪造痕迹。额外引入的频率模态作为RGB信息的补充,能够提高模型对各种图像压缩算法^[20]的鲁棒性。基于以上问题,本文提出基于RGB图像和频域特征的双分支编码器,并结合噪声自增强和相互增强的双流增强模

块,实现多种篡改线索的充分交互。

2.2 渐进整合

神经网络中不同阶段的特征是互补的,深层特征包含全局信息,更有可能定位被篡改的区域,而浅层特征提供更多的空间细节。得益于多尺度特征的逐步集成,多级网络在计算机视觉领域已广泛应用于像素级任务。Li等人^[21]逐步整合多阶段特征映射来处理复杂场景下的图像去雾。Chen等人^[22]观察到的是跨阶段的互补信息,这在以往的工作中没有充分考虑到。他们提出一种渐进式互补感知融合网络用于RGB-D显著目标检测。Liu等人^[23]利用多尺度特征以粗到细的方式产生图像篡改定位掩模,证明高阶的信息可以细化低阶段的特征。

因此,本文自适应地重新校准不同阶段的定位特征。首先通过探索长程语义依赖来定位潜在的篡改区域,然后通过在每个解码模块中注入深层语义特征和预测以及主干当前级别特征对解码器进行重构,以逐步细化粗略预测。

3 网络结构

本文提出基于双流增强编码和注意优化解码的图像篡改定位算法,如图1所示。首先,利用RGB特征流和频率特征流分别提取图像内容和高频信息。其次,提出双流增强编码对图像和频域特征进行基于噪声和通道注意力的自增强和基于特征映射的交叉注意权重的交互增强。然后,邻阶特征聚合模块对经过多尺度感受野的增强特征进行邻阶聚合,实现空间信息和语义信息的相互补充。最后,注意优化解码模块利用篡改区域和非篡改区域的协同作用,逐步优化预测的篡改区域。

3.1 双流特征编码

篡改图像通常经过后处理以隐藏篡改伪影,造成单一图像流错误地抑制篡改内容信息,因此,本定位提供充分的可表征特征。图像流旨在提取视觉上可感知的篡改伪影,包括不自然的篡改区域边界、篡改区域与真实区域间对比度差异等特征。本文采用ImageNet数据集上预训练的ResNet50作为骨干网络对输入图像 $X \in R^{H \times W \times 3}$ 提取特征,标记为 $R_i, i = \{1, 2, 3, 4, 5\}$ 。

本文对RGB图像 X 进行离散余弦变换

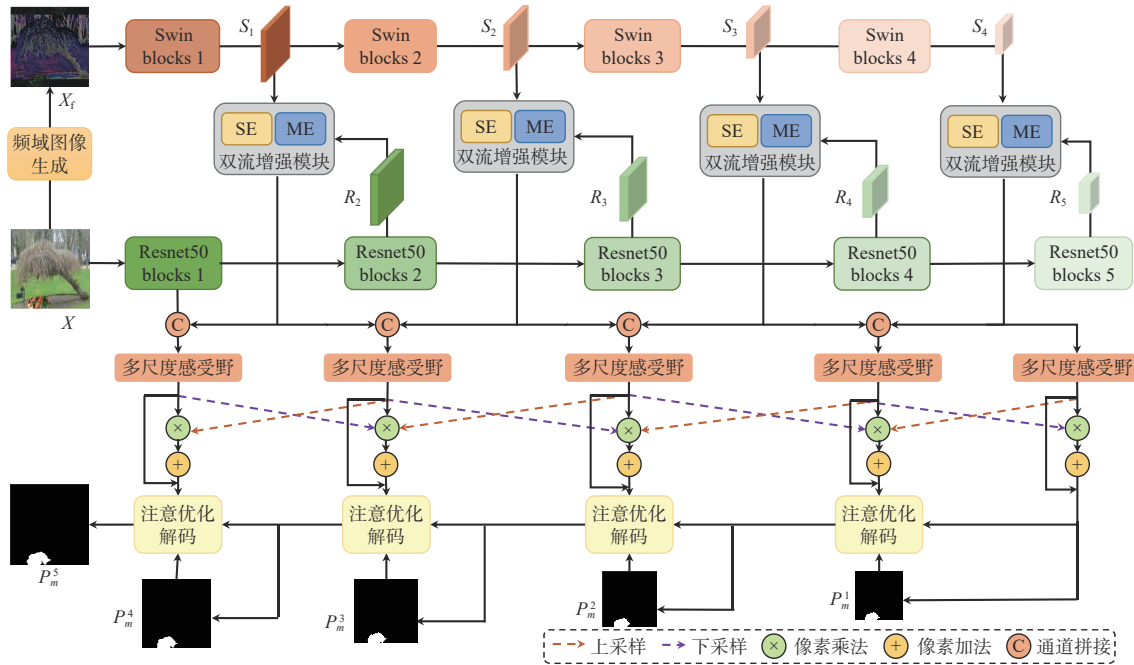


图1 基于双流增强编码和注意优化解码的图像篡改定位网络架构

Fig. 1 Network of dual-stream enhancement encoder and attention optimization decoder for image manipulation

(Discrete Cosine Transform, DCT), 将其从 RGB 域转换到频域 $X_q \in R^{H \times W \times 1}$ 。为去除噪声, 本文采用高通滤波器^[24]获取高频分量, 并将其逆变换回 RGB 域得到频域图像 X_f , 以保持图像的移位不变性和局部一致性, 如式(1)所示:

$$X_f = D^{-1}(F(X_q, \alpha)), \quad (1)$$

式中: F 表示高通滤波器, α 是手动设计的阈值以控制低频成分被过滤掉, D^{-1} 表示 DCT 逆变换。为能够为篡改定位提供充足的线索, 本文使用 Swin transformer 作为频率编码分支的特征提取骨干, 将 X_f 输入到 Swin transformer 中, 提取全局高

频特征和局部高频特征, 标记为 $S_i, i = \{1, 2, 3, 4\}$ 。由于 Swin transformer 特征提取时先对图像进行分块, 因此 S_i 的尺度与 R_{i+1} 一致。

3.2 双流增强模块

当前双流特征大都采用池化层或通道拼接操作进行融合交互, 但该类方法融合互补不充分。因此, 本文提出双流增强模块 (Dual-stream Enhancement Model, DEM), 由自增强模块和交互增强模块组成, 旨在增强图像流和频率流的相同尺度的特征并交互补充增强, 更加专注于篡改特征, 如图 2 所示。

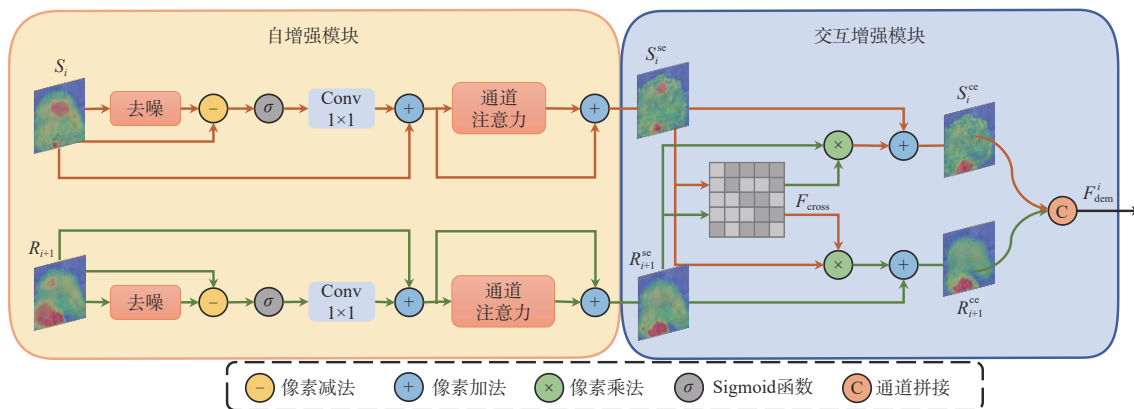


图2 双流增强模块

Fig. 2 Framework of dual-stream enhancement model

3.2.1 双流自增强

本模块由两个子模块组成:噪声增强和通道注意,分别对每个流的特征进行增强。近年来,篡改手段愈发先进,将细微噪声引入到图像篡改定位任务发挥了更关键的作用。基于此,本文设计了噪声增强模块来挖掘隐藏在特征噪声中的伪影,特别是高频特征中的伪影。

为进行噪声增强,我们首先提取输入图像流 R_{i+1} 和频域流 S_i 特征图的噪声,应用 Sigmoid 函数对微小噪声进行放大,抑制噪声的夸大值以防止极化。最后,通过输入特征减去空间去噪特征得到噪声信息 S_i^{ne} ,如式(2)所示:

$$S_i^{ne} = S_i + \text{Conv}(\sigma(S_i - M(S_i))), \quad (2)$$

式中: M 表示内核大小为 3×3 的中值滤波器,可过滤椒盐等噪声进行特征级去噪; $\sigma(\cdot)$ 表示 Sigmoid 函数。我们进一步利用通道注意块促进每个流中的通道间交互,即结合所有通道信息确定每个通道的重要性权重,得到自增强特征 S_i^{se} :

$$S_i^{se} = S_i^{ne} + S_i^{ne} \otimes \sigma(\text{MLP}(\text{GAP}(S_i^{ne}) + \text{GMP}(S_i^{ne}))), \quad (3)$$

式中: MLP 为多层感知机, GAP 为全局平均池化, GMP 为全局最大池化。

3.2.2 双流交互增强

由于两个流的特征不同,它们提供了不同的线索,可以相互补充。因此,本文设计了相互增强模块。不同于以往方法中广泛使用的简单通道拼接,本文首先计算两流特征映射的交叉注意权重,采用交叉注意矩阵将注意特征从一个流扩展到另一个流。具体来说,首先对经过自增强模块的两流特征 S_i^{se} 和 R_{i+1}^{se} 使用交叉注意力得到交叉注意矩阵 F_{cross} :

$$F_{\text{cross}} = \text{soft max} \left\{ \text{Cat} \left(R_{i+1}^{se} W_{eq}, S_i^{se} W'_{eq} \right) \cdot \left[\text{Cat} \left(R_{i+1}^{se} W_{ek}, S_i^{se} W'_{ek} \right)^T \right] \right\}, \quad (4)$$

式中: W_{eq} 、 W'_{eq} 、 W_{ek} 、 W'_{ek} 表示4个线性投影层的可学习参数; S_i^{se} 和 R_{i+1}^{se} 分别与交叉注意矩阵 F_{cross} 相乘,再相加得到交互增强特征 S_i^{ce} 和 R_{i+1}^{ce} ,进行拼接后得到双流增强特征 F_{dem}^i 。

3.3 邻阶特征聚合

不同尺度特征融合在语义信息和空间信息的结合中起着至关重要的作用。浅层低级特征缺乏语义信息,但保留了丰富的空间细节。低级特征的噪声较高而高级特征缺乏空间信息。因

此,本文提出将相邻特征进行融合,以抑制来自浅层的特征噪声,同时增强深层特征的代表能力。

为模拟人类视觉系统并在搜索过程中包含更多样化的特征表示,本文提出感受野模块以探索多尺度上下文信息。具体来说,将相邻的增强特征 F_{dem}^i 拼接后作为多尺度感受野的输入,其中为了充分利用浅层空间信息和深层语义信息,将 F_{dem}^1 与 R_1 拼接后的特征和 F_{dem}^4 分别作为输入。输入特征图依次经过3个分别由 $1 \times (2k-1)$ 卷积层构成的分支 $\{b_k, k=3, 4, 5\}$ 。将3个分支通道拼接后经过 1×1 卷积层和残差连接与输入特征拼接后,得到特征 F_{rf}^i 。

本文提出邻阶特征聚合模块进行空间信息和语义信息的相互补充,通过邻近层的特征相乘获得重要语义信息,与输入特征 F_{rf}^i 进行相加,保留特征的细节信息,增强重要语义信息,得到聚合特征 F_{plam}^i :

$$F_{\text{plam}}^i = F_{\text{rf}}^i + F_{\text{rf}}^{i-1} \otimes F_{\text{rf}}^i \otimes F_{\text{rf}}^{i+1}. \quad (5)$$

3.4 注意优化解码

通过研究模拟人类发现篡改异常区域过程,发现人类通常利用视觉上的不一致反复对比篡改区域和真实区域。基于此,本文利用篡改区域和非篡改区域协同增强模型的篡改定位能力,提出注意优化解码模块,消除初始篡改区域预测中边缘像素的错误预测,逐步精确细化篡改定位。受文献[23]的启发,本文提出注意优化解码模块(Attention Optimization Decoder, AOD),如图3所示。

注意优化解码模块中首先对更高层次的预测进行上采样,并用 Sigmoid 层对其进行归一化。然后对归一化映射 P_m^i 和其取反映射 $\bar{P}_m^i$ 分别乘以当前级聚合特征 F_{plam}^{i+1} ,再分别经过一系列卷积操作,分别发现错误的篡改像素预测 F_{fmp} 和错误的非篡改像素预测 F_{fnmp} ,最后将 F_{fmp} 和 F_{fnmp} 与高层解码特征 F_{aod}^i 进行逐像素加减操作来优化错误的像素预测,优化方法如式(6)和式(7)所示:

$$F_{\text{up}} = \text{Upsample}(\text{BRC}(F_{\text{aod}}^i)), \quad (6)$$

$$F_{\text{aod}}^{i+1} = \text{BR}(F_{\text{fmp}} + \text{BR}(F_{\text{up}} - F_{\text{fnmp}})), \quad (7)$$

式中: F_{aod}^i 和 F_{aod}^{i+1} 分别表示输入的高级特征和输出的优化特征; BRC 是卷积层、批处理归一化(Batch Normalization, BN)和 ReLU 的结合; BR 是批处理归一化和 ReLU 的结合; Upsample

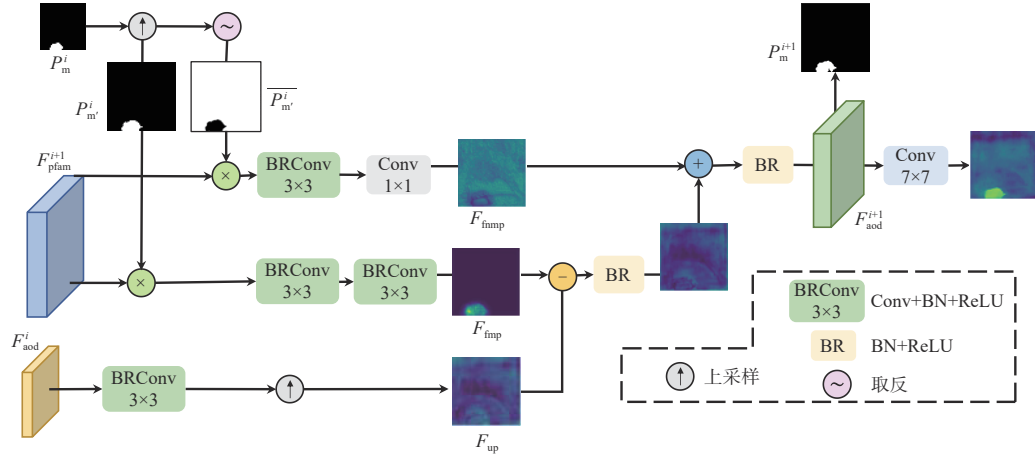


图3 注意优化解码模块

Fig. 3 Model of attention optimization decoder

为双线性上采样。本文使用逐元素的减法操作来抑制篡改像素的干扰,并进行逐元素的加法增强缺失的篡改像素。最后,在优化后的特征 F_{aom}^{i+1} 上应用卷积层,以得到更精确的预测图,使用真实值来监督生成的预测图。

3.5 损失函数

图像篡改定位任务中,像素的错误预测、篡改区域定位偏差以及篡改区域边界预测模糊等问题会损失位置信息和结构信息。篡改图像中篡改像素比例相对较小,会导致类不平衡问题,使模型倾向于将像素分类为真实^[26]。基于此,本文采用多尺度深度混合损失函数 L 进行监督^[27],包括二元交叉熵(BCE)损失^[25]、交并比(IoU)损失^[28]和结构相似度(SSIM)损失^[29],如式(8)所示:

$$L = \sum_{k=1}^5 L_k^{\text{BCE}} + L_k^{\text{SSIM}} + L_k^{\text{IoU}}, \quad (8)$$

式中, $k \in \{1, 2, 3, 4, 5\}$ 表示第 k 个尺度的监督输出。

二进制交叉熵(BCE)对前景像素和背景像素一视同仁。BCE损失 L^{BCE} 定义为:

$$L^{\text{BCE}} = - \sum_j [G_j \log P_j + (1 - G_j) \log (1 - P_j)], \quad (9)$$

式中: P 表示预测的特征映射, G 表示真实值, j 表示像素坐标。

SSIM损失在图像块上衡量预测和真实值之间的结构相似性,赋予篡改区域和真实区域间边界像素更高的权重,以保留边界区域的结构信息。SSIM损失 L^{SSIM} 的计算如式(10)所示:

$$L^{\text{SSIM}} = 1 - \frac{(2\mu_P\mu_G + \tau_1)(2\sigma_{PG} + \tau_2)}{(\mu_P^2 + \mu_G^2 + \tau_1)(\sigma_P^2 + \sigma_G^2 + \tau_2)}, \quad (10)$$

式中, μ_P 、 μ_G 和 σ_P 、 σ_G 分别表示预测结果 P 和真实值 G 的均值和标准差; σ_{PG} 表示协方差矩阵; $\tau_1 = 0.01^2$ 和 $\tau_2 = 0.03^2$ 都是常数,以防被0整除。

IoU损失 L^{IoU} 经常被用于目标检测与分割,强调更大的前景区域,如式(11)所示:

$$L^{\text{IoU}} = 1 - \sum_j \frac{P_j G_j}{P_j + G_j - P_j G_j}. \quad (11)$$

4 实验

4.1 实验设置

4.1.1 数据集

为与主流方法进行全面比较,本文在DEFACTO^[30]数据集上进行训练,在4个主流公共基准数据集:NIST16^[31]、Coverage^[32]、Columbia^[33]、CASIA^[34]和两个现实挑战数据集:IMD20^[35]、Wild^[36]上进行测试。DEFACTO是一个大规模数据集,包含149k幅由MSCOCO^[37]通过复制-粘贴、拼接和移除操作自动生成的篡改图像。基于本文提出的模型参数量,本文随机选择60k幅篡改图像作为训练和测试数据,训练测试比为9:1。在公开数据集NIST16、Coverage和CASIA上,本文遵循与SPAN^[11]和PSCC-Net^[23]一致的数据集训练-测试比率划分。为验证模型的泛化性,本文在Columbia、IMD20和Wild数据集上不进行微调,仅进行测试。详细的数据集信息如表1所示。

表 1 图像数据集细节

Tab. 1 Details of image datasets

数据集	训练集	测试集	复制-粘贴	拼接	移除	后处理
DEFACTO ^[30]	54 000	6 000	✓	✓	✓	✓
NIST16 ^[31]	404	160	✓	✓	✓	✓
Coverage ^[32]	75	25	✓	×	×	✓
CASIA ^[34]	5 123(v2.0)	920(v1.0)	✓	✓	×	✓
Columbia ^[33]	0	180	×	✓	×	×
IMD20 ^[35]	0	2 010	✓	✓	✓	✓
Wild ^[36]	0	201	×	✓	×	×

4.1.2 评价指标

为全面定量分析模型的定位性能,本文采用多种度量指标,即F1分数、交并比IoU和ROC曲线下的面积AUC进行评价。F1和IoU的计算如式(12)和式(13)所示:

$$F1 = \frac{2*TP}{2*TP + FP + FN}, \quad (12)$$

$$IoU = \frac{TP}{TP + FP + FN}, \quad (13)$$

式中:TP表示检测正确的篡改像素数量,FP和FN分别表示预测错误的篡改像素和真实像素数量,阈值为默认值0.5。

4.1.3 实现细节

本文基于PyTorch实现,并在GeForce GTX 3060上训练,批处理大小为16。由于计算设备的内存限制,本文遵循Mantra-net处理方式,将所有输入图像的大小调整为 256×256 。为节省计算资源和减少运行时间,使用在ImageNet数据集上的预训练权重对两流网络进行初始化,Swin transformer块的窗口和补丁大小分别为8和2。采用Adam优化器($\beta_1 = 0.9, \beta_2 = 0.999$,

学习率 $\alpha = 10^{-4}$,权值衰减 $\beta = 10^{-4}$)优化网络参数,采用多步衰减策略更新新学习率 $\gamma = 0.5$ 。总训练周期为50个epoch。

4.2 消融实验

4.2.1 模型消融

为了验证模型中各个模块的有效性,本文在DEFACTO数据集进行消融实验,定量结果如表2所示。所有模型都采用相同的设置进行训练,实验结果证明,本文模型的各个模块有效,在F1、IoU和AUC的各项评价指标都有显著提高。Baseline+DEM(w/o ME)通过引入细微噪声来提高F1、IoU和AUC评价指标,验证DEM通过引入噪声信息来增强模型关注篡改区域的能力。加入交互增强模块后的结果表明,两流特征是不同和互补的,交互增强模块可增强篡改特征。Baseline+DEM+RF通过加入感受野模块提升AUC、F1和IoU,提高模型在大范围内感知丰富上下文的能力。在此基础上引入邻阶特征聚合模块(PFAM)后,AUC、F1和IoU分别提高0.9%、1.6%和3.1%,表明PFAM模块通过融合不同尺度特征,

表 2 DEFACTO数据集上的消融实验定量结果

Tab. 2 Quantitative results of ablation experiments on DEFACTO dataset

模型	组成模块					AUC	F1	IoU
	自增强	交互增强	感受野	邻阶聚合	注意优化			
Baseline						0.961	0.875	0.778
Baseline+DEM(w/o ME)	✓					0.969	0.894	0.808
Baseline+DEM	✓	✓				0.978	0.911	0.837
Baseline+DEM+RF	✓	✓	✓			0.982	0.921	0.853
Baseline+DEM+RF+PFAM	✓	✓	✓	✓		0.988	0.936	0.880
Baseline+DEM+RF+PFAM+AOD(w/o FMP)	✓	✓	✓	✓	✓	0.989	0.941	0.892
Baseline+DEM+RF+PFAM+AOD	✓	✓	✓	✓	✓	0.993	0.949	0.907

注:加粗数据为最优值

可抑制来自浅层的特征噪声,同时增强深层特征的表示能力。本文算法从背景、前景以及全局角度渐进细化篡改特征,并逐步聚焦篡改区域,F1、IoU和AUC分别提高了1.3%、3.0%和0.5%。

为验证消融实验的可视化结果,本文在NIST16数据集上进行可视化,如图4所示。由第三列可以看出,虽然Baseline+DEM(w/o ME)模型可以基本完成对拼接篡改图像的整体篡改定位,但定位结果仍比较粗糙,丢失了较多结构

细节。而对于复制-粘贴和移除篡改图像,Baseline+DEM(w/o ME)模型的误检区域较大,性能较差。增加交互增强模块后,Baseline+DEM模型对于复制-粘贴和移除篡改图像的定位结果更加精确。邻阶特征聚合模块对复制-粘贴和移除篡改图像可基本定位篡改区域,并削弱了在拼接篡改图像上的像素误检。最终,在注意优化解码模块的细化下篡改区域更加精细准确,证明了其细化特征的有效性。

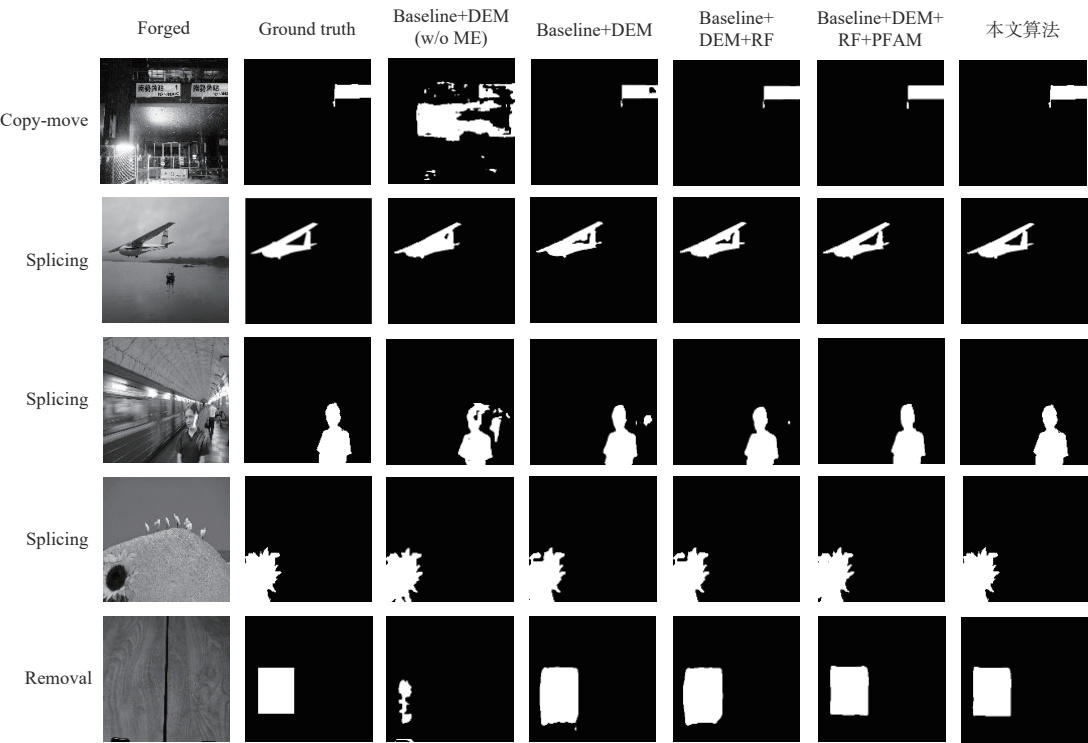


图4 NIST数据集消融实验的可视化结果

Fig. 4 Qualitative visualization results of ablation experiments on NIST dataset

4.2.2 去噪方法

为探索不同类型、核大小和组合的滤波器,寻找合适的去噪方法,本文在DEFACTO数据集上进行消融实验,结果如表3所示。本文使用的中值滤波器优于均值滤波器、非局部均值滤波器和高斯滤波器。对于内核大小,分别设置为3、5、7,结果发现最小的内核性能最好。我们进一步组合多个中值滤波器来提取不同大小区域的噪声,采用1×1卷积层进行通道维度降维。结果表明,单个中值滤波器比多个滤波器更好。基于以上分析,选择核大小为3的单个中值滤波器作为本文方法的去噪设置。

表3 DEFACTO数据集上去噪方法的定量结果

Tab.3 Quantitative results of denoising methods on DEFACTO dataset

去噪方法	AUC
中值滤波(ks=3)	0.993
均值滤波(ks=3)	0.983
非局部均值滤波(ks=3)	0.979
高斯滤波(ks=3)	0.982
中值滤波(ks=5)	0.985
中值滤波(ks=7)	0.988
多个中值滤波(ks=3,5)	0.987
多个中值滤波(ks=3,5,7)	0.986

注:加粗数据为最优值;ks为内核大小

4.3 对比实验

4.3.1 预训练模型定量实验

对于预训练模型的评估,我们将模型与 ManTra-Net^[10]、SPAN^[11]、PSCC-Net^[23]、TBNe^[38]、SSDA-Ne^[39]、ForensicsOSN^[40]、MVSS-Net++^[41]

和 ObjectFormer^[12] 进行比较,其中 ManTra-Net、PSCC-Ne 和 MVSS-Net++ 的实验结果是利用论文开源代码在本文训练数据集训练;TBNet 和 ObjectFormer 没有开源代码,实验结果源于作者公开发表的论文,如表 4 所示。

表 4 预训练模型的定量比较

Tab. 4 Quantitative comparison of pre-trained models

方法	Columbia		IMD20		Wild	
	AUC	F1	AUC	F1	AUC	F1
ManTra-Net ^[10]	0.824	0.704	0.758	0.675	0.654	0.435
SPAN ^[11]	0.936	0.815	—	—	—	—
PSCC-Net ^[23]	0.982	0.849	0.812	0.746	0.745	0.529
TBNet ^[38]	—	—	—	0.751	—	0.567
SSDA-Net ^[39]	—	—	—	—	—	0.418
ForensicsOSN ^[40]	0.862	0.707	—	—	—	—
MVSS-Net++ ^[41]	0.804	0.744	0.806	0.758	0.762	0.557
ObjectFormer ^[12]	0.955	—	0.821	—	—	—
本文算法	0.988	0.893	0.853	0.810	0.812	0.588

注:加粗数据为最优值

本文算法在 Columbia、IMD20 和 Wild 数据集上的定位性能最好。本文方法的 AUC 性能较次优的 ObjectFormer 在 Columbia 数据集上提高了 3.5%。在 IMD20 和 Wild 数据集上,本文算法的 AUC 较 MVSS-Net++ 分别提高 5.8% 和 6.6%,表明本文方法具有较强的捕获篡改特征的能力,可以很好地推广到各种高质量的被篡改

图像数据集。

4.3.2 微调模型定量实验

我们使用训练策略在特定的数据集上进一步微调预训练模型,然后通过对训练数据的交叉验证在每个测试数据集上选择最佳的微调模型。在微调设置下,像素级 AUC 和 F1 分数的定量比较如表 5 所示。本文算法在 Coverage 数据集上

表 5 微调模型的定量对比实验

Tab. 5 Quantitative comparison of fine-tuning models

方法	测试方式	NIST16		Coverage		CASIA	
		AUC	F1	AUC	F1	AUC	F1
ELA ^[42]	无监督	0.429	0.236	0.583	0.222	0.613	0.214
NOI ^[43]	无监督	0.487	0.285	0.587	0.269	0.612	0.263
CFA1 ^[4]	无监督	0.501	0.174	0.485	0.190	0.522	0.207
J-LSTM ^[44]	微调	0.764	—	0.614	—	—	—
RGB-N ^[9]	微调	0.937	0.722	0.817	0.437	0.795	0.408
H-LSTM ^[13]	微调	0.794	—	0.712	—	—	—
GSR-Net ^[20]	微调	0.945	0.736	0.768	0.489	0.796	0.574
SPAN ^[11]	微调	0.961	0.582	0.937	0.558	0.838	0.382
PSCC-Net ^[23]	微调	0.996	0.819	0.941	0.723	0.875	0.554
MVSS-Net++ ^[41]	微调	0.979	0.872	0.903	0.752	0.836	0.551
ObjectFormer ^[12]	微调	0.996	0.824	0.957	0.758	0.882	0.579
本文算法	微调	0.997	0.912	0.974	0.788	0.898	0.624

注:加粗数据为最优值

的 AUC 比 ObjectFormer 高出 1.8%，在 NIST16 上的 F1 高出 10.6%。从表 5 可以清楚地看出，本文模型在 DEFACTO 数据集上进行了训练，并在每个基准数据集上进行了微调，达到了最先进的性能，这表明所提出的模块之间能够协同作用，以捕获细微的篡改痕迹。

4.3.3 对比方法的定性实验

图 5 显示了本文方法与其他最先进方法在 4 个主流公共基准数据集和两个现实挑战数据集上的定性结果比较。这些结果表明，本文方法不仅在公

共基准数据集上分割的篡改区域结果准确精细，而且在现实挑战数据集上也能对篡改区域准确定位。尽管 MVSS-Net++ 能够粗略定位出某些篡改区域，但其定位的结果比较粗糙并且出现了很多误检区域。反观本文方法在各个数据集上都能够准确定位各种类型的篡改区域。此外，本文方法还能够准确定位小目标篡改区域。例如，在第 7 列中本文方法能够准确定位比较小的复制-粘贴篡改区域。这些定性可视化结果表明，本文方法可以捕捉细微的篡改痕迹，并产生更准确精细的定位结果。

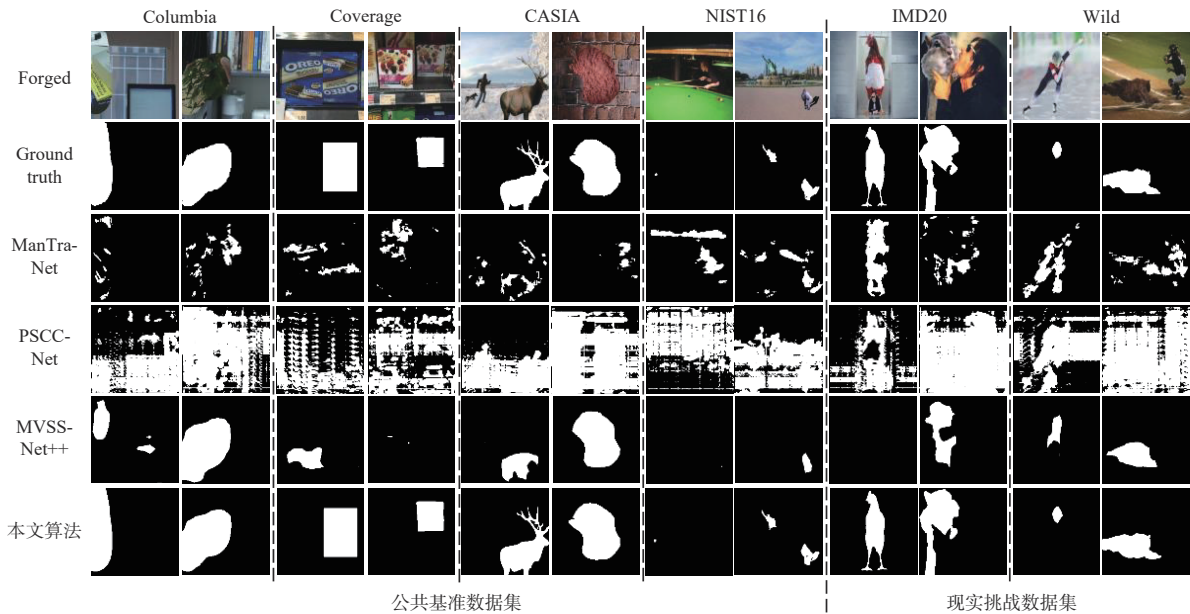


图 5 4 个公共基准数据集和两个现实挑战数据集上的定性对比结果

Fig. 5 Qualitative comparison results on four public benchmark datasets and two real-world challenge datasets

4.4 鲁棒性实验

我们在 NIST16 数据集的篡改图像上应用不同的图像后处理方法来评估本文模型的鲁棒性，包括不同尺度的图像缩放、核大小为 k 的高斯模

糊、标准差为 σ 的高斯噪声和质量系数为 q 的 JPEG 压缩。将本文方法的预训练模型与 ManTra-Net、SPAN 以及 PSCC-Net 在这些失真图像上进行比较，结果如图 6 所示。本文模型对各种

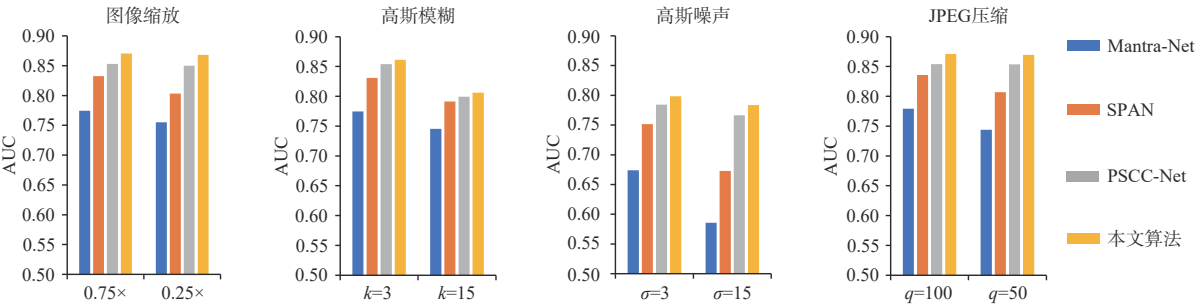


图 6 NIST16 数据集上的鲁棒性实验

Fig. 6 Robustness experiment on the NIST16 dataset

失真技术表现出更好的鲁棒性。

5 结 论

本文提出基于双流增强编码和注意优化解码的图像篡改定位框架。双流增强编码分别对图像和频域特征进行噪声和通道注意力的自增强和特征映射的交叉注意权重的交互增强,并利用基于篡改区域和非篡改区域协同增强的注意优化解码模块消除初始篡改区域预测中边缘像素的错误预测,逐步精确细化篡改定位。本文在4个主流公共基准数据集和两个现实挑战数据集上与主流篡改定位方法进行了充分的对比实验。本文算法的无微调模型在IMD20和Wild数据集上,AUC较MVSS-Net++方法分别提高5.8%

和6.6%。微调模型在Coverage数据集上的AUC和NIST16数据集上的F1分别较ObjectFormer方法高出1.8%和10.6%。实验结果表明,本文提出的篡改定位框架能够充分利用多种篡改操作特征,在6个主流篡改数据集上实现篡改区域的精确定位,具有更高的定位精度和更强的鲁棒性。

但是在编码过程中,本文使用ResNet50和Transformer同时提取输入图像的深度特征,因此不可避免地存在冗余特征。除此之外,每幅篡改图像上仅只使用了一种篡改手段。我们未来将重构网络以适应多操作方法的混合情况,并设计特征选择策略,以避免过多的冗余特征传递到网络中。并且,将研究一幅篡改图像包含多类篡改操作的取证任务,这仍然是一个开放的研究挑战。

参 考 文 献:

- [1] 张玉林,王宏霞,张瑞,等. 语义一致性引导的多任务拼接篡改检测[J]. 中国图象图形学报,2023,28(3):775-788.
ZHANG Y L, WANG H X, ZHANG R, *et al.* Semantic consistency-relevant multitask splicing-tampered detection [J]. *Journal of Image and Graphics*, 2023, 28(3): 775-788. (in Chinese)
- [2] 林晶,黄添强,林玲鹏,等. 采用局部强度顺序模式的图像复制-粘贴篡改检测算法[J]. 通信学报,2016,37(S1): 132-139.
LIN J, HUANG T Q, LIN L P, *et al.* Detection of image copy-move forgery using local intensity order pattern [J]. *Journal on Communications*, 2016, 37(S1): 132-139. (in Chinese)
- [3] ZHU X S, QIAN Y J, ZHAO X F, *et al.* A deep learning approach to patch-based image inpainting forensics [J]. *Signal Processing: Image Communication*, 2018, 67: 90-99.
- [4] FERRARA P, BIANCHI T, DE ROSA A, *et al.* Image forgery localization *via* fine-grained analysis of CFA artifacts [J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(5): 1566-1577.
- [5] LIU B, PUN C M, YUAN X C. Digital image forgery detection using JPEG features and local noise discrepancies [J]. *The Scientific World Journal*, 2014, 2014: 230425.
- [6] IAKOVIDOU C, ZAMPOGLOU M, PAPADOPOULOS S, *et al.* Content-aware detection of JPEG grid inconsistencies for intuitive image forensics [J]. *Journal of Visual Communication and Image Representation*, 2018, 54: 155-170.
- [7] PANG Y W, ZHAO X Q, ZHANG L H, *et al.* Multi-scale interactive network for salient object detection [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2020: 9410-9419.
- [8] QIU Y, LIU Y, CHEN Y N, *et al.* A2SPPNet: attentive atrous spatial pyramid pooling network for salient object detection [J]. *IEEE Transactions on Multimedia*, 2022, 25: 1991-2006.
- [9] ZHOU P, HAN X T, MORARIU V I, *et al.* Learning rich features for image manipulation detection [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 1053-1061.
- [10] WU Y, ABDALMAGEED W, NATARAJAN P. ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA: IEEE, 2019: 9535-9544.

- [11] HU X F, ZHANG Z H, JIANG Z Y, *et al.* SPAN: Spatial pyramid attention network for image manipulation localization [C]// *Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: IEEE, 2020: 312-328.
- [12] WANG J K, WU Z X, CHEN J J, *et al.* ObjectFormer for image manipulation detection and localization [C]// *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, LA, USA: IEEE, 2022: 2354-2363.
- [13] BAPPY J H, SIMONS C, NATARAJ L, *et al.* Hybrid LSTM and encoder-decoder architecture for detection of image forgeries [J]. *IEEE Transactions on Image Processing*, 2019, 28(7): 3286-3300.
- [14] CHEN H P, CHANG C Q, SHI Z N, *et al.* Hybrid features and semantic reinforcement network for image forgery detection [J]. *Multimedia Systems*, 2022, 28(2): 363-374.
- [15] LI S B, XU S B, MA W, *et al.* Image manipulation localization using attentional cross-domain CNN features [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(9): 5614-5628.
- [16] LIN X, WANG S, DENG J H, *et al.* Image manipulation detection by multiple tampering traces and edge artifact enhancement [J]. *Pattern Recognition*, 2023, 133: 109026.
- [17] ZHUO L, TAN S Q, LI B, *et al.* Self-adversarial training incorporating forgery attention for image forgery localization [J]. *IEEE Transactions on Information Forensics and Security*, 2022, 17: 819-834.
- [18] CHEN X R, DONG C B, JI J Q, *et al.* Image manipulation detection by multi-view multi-scale supervision [C]// *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, QC, Canada: IEEE, 2021: 14165-14173.
- [19] WANG J K, WU Z X, OUYANG W H, *et al.* M2TR: Multi-modal multi-scale transformers for deep fake detection [C]// *Proceedings of the 2022 International Conference on Multimedia Retrieval*. Newark, NJ, USA: IEEE, 2022: 615-623.
- [20] ZHOU P, CHEN B C, HAN X T, *et al.* Generate, segment, and refine: Towards generic manipulation segmentation [C]// *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*. New York: AAAI Press, 2020: 13058-13065.
- [21] 乔丹, 张闯, 朱晨雨. 基于多尺度融合卷积神经网络的图像去雾算法[J]. 液晶与显示, 2021, 36(10): 1420-1429.
QIAO D, ZHANG C, ZHU C Y. Image dehazing algorithm based on multi-scale concat convolutional neural network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(10): 1420-1429. (in Chinese)
- [22] CHEN H, LI Y F. Progressively complementarity-aware fusion network for RGB-D salient object detection [C]// *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 3051-3060.
- [23] LIU X H, LIU Y J, CHEN J, *et al.* PSCC-Net: Progressive spatio-channel correlation network for image manipulation detection and localization [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(11): 7505-7517.
- [24] 刘玉洁, 朱韶平. 基于全局和局部多特征的图像增强算法[J]. 液晶与显示, 2020, 35(5): 508-512.
LIU Y J, ZHU S P. Image enhancement algorithm based on global and local multi features [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(5): 508-512. (in Chinese)
- [25] SHORE J, JOHNSON R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy [J]. *IEEE Transactions on Information Theory*, 1980, 26(1): 26-37.
- [26] ZHANG Y L, ZHU G P, WU L G, *et al.* Multi-task SE-network for image splicing localization [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4828-4840.
- [27] 陆康亮, 薛俊, 陶重彝. 融合空间掩膜预测与点云投影的多目标跟踪[J]. 光电工程, 2022, 49(9): 220024.
LU K L, XUE J, TAO C B. Multi target tracking based on spatial mask prediction and point cloud projection [J]. *Opto-Electronic Engineering*, 2022, 49(9): 220024. (in Chinese)
- [28] YU J H, JIANG Y N, WANG Z Y, *et al.* UnitBox: An advanced object detection network [C]// *Proceedings of the 24th ACM international conference on Multimedia*. Amsterdam: IEEE, 2016: 516-520.
- [29] 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.

- [30] MAHFOUDI G, TAJINI B, RETRAINT F, *et al.* DEFACTO: Image and face manipulation dataset [C]. 2019 27th European Signal Processing Conference (EUSIPCO). A Coruna, Spain: IEEE, 2019: 1-5.
- [31] GUAN H Y, KOZAK M, ROBERTSON E, *et al.* MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation [C]. 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). Waikoloa, HI, USA: IEEE, 2019: 63-72.
- [32] WEN B H, ZHU Y, SUBRAMANIAN R, *et al.* COVERAGE—A novel database for copy-move forgery detection [C]. 2016 IEEE International Conference on Image Processing (ICIP). Phoenix, AZ, USA: IEEE, 2016: 161-165.
- [33] NG T T, HSU J, CHANG S F. Columbia image splicing detection evaluation dataset [R]. *DVMM Lab. Columbia Univ CalPhotos Digit Libr*, 2009.
- [34] DONG J, WANG W, TAN T N. CASIA image tampering detection evaluation database [C]. 2013 IEEE China Summit and International Conference on Signal and Information Processing. Beijing, China: IEEE, 2013: 422-426.
- [35] NOVOZÁMSKÝ A, MAHDIAN B, SAIC S. IMD2020: A large-scale annotated dataset tailored for detecting manipulated images [C]//*Proceedings of the 2020 IEEE/CVF Winter Applications of Computer Vision Workshops*. Snowmass, CO, USA: IEEE, 2020: 71-80.
- [36] HUH M, LIU A, OWENS A, *et al.* Fighting fake news: Image splice detection via learned self-consistency [C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer, 2018: 106-124.
- [37] LIN T Y, MAIRE M, BELONGIE S, *et al.* Microsoft COCO: Common objects in context [C]. 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 740-755.
- [38] GAO Z, SUN C, CHENG Z, *et al.* TBNet: A two-stream boundary-aware network for generic image manipulation localization [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(7): 7541-7556.
- [39] RAO Y, NI J Q, ZHANG W Z, *et al.* Towards jpeg-resistant image forgery detection and localization via self-supervised domain adaptation [J/OL]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022: 1-12. <https://ieeexplore.ieee.org/document/9904872>.
- [40] WU H W, ZHOU J T, TIAN J Y, *et al.* Robust image forgery detection against transmission over online social networks [J]. *IEEE Transactions on Information Forensics and Security*, 2022, 17: 443-456.
- [41] DONG C B, CHEN X R, HU R H, *et al.* MVSS-Net: Multi-view multi-scale supervised networks for image manipulation detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(3): 3539-3553.
- [42] KRAWETZ N, SOLUTIONS H F. A picture's worth [J]. *Hacker Factor Solutions*, 2007, 6(2): 2.
- [43] MAHDIAN B, SAIC S. Using noise inconsistencies for blind image forensics [J]. *Image and vision computing*, 2009, 27(10): 1497-1503.
- [44] BAPPY J H, ROY-CHOWDHURY A K, BUNK J, *et al.* Exploiting spatial structure for localizing manipulated image regions [C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE, 2017: 4980-4989.

作者简介:



朱 叶,女,博士,副教授,2017年于吉林大学获得博士学位,主要从事图像处理与模式识别以及多媒体安全方面的研究。E-mail:zhuye@hebut.edu.cn



于 洋,男,博士,副教授,2012年于河北工业大学获得博士学位,主要从事图像处理与模式识别方面的研究。E-mail:yuyang@hebut.edu.cn