基于边界不确定性学习的图像篡改定位方法

陈海鹏[1,2],刘宏昕[1,2],康 辉[1,2],刘雪洁[1]

(1.吉林大学 计算机科学与技术学院,长春 130012;2.吉林大学 符号计算与知识工程教育部重点实验室,长春 130012)

摘 要:针对目前图像篡改定位方法提取特征尺度单一、小篡改区域易与背景混淆造成误检漏检、预测结果不确定性高等问题,提出了基于边界不确定性学习的图像篡改定位方法。首先,使用金字塔视觉变压器提取篡改图像基础特征。其次,利用多级交互粗定位分支生成粗定位图。再次,利用小目标感知细化分支提高小篡改区域感知定位能力。随后,利用多尺度特征融合模块实现多尺度特征的充分交互与融合。最后,提出基于熵的边界不确定性感知损失进行辅助监督,极大地降低了预测结果的不确定性。在5个常用公开图像篡改数据集上分别进行域内和跨域实验,结果表明,本文方法可精准定位篡改区域,并优于其他方法。

关键词:计算机应用;图像篡改定位;多尺度特征交互;小篡改区域感知细化;边界不确定性学习

中图分类号:TP391 **文献标志码**:A **文章编号**:1671-5497(2025)12-4063-09

DOI:10.13229/j.cnki.jdxbgxb.20250014

Image manipulation localization method based on boundary uncertainty learning

CHEN Hai-peng[1,2],LIU Hong-xin[1,2],KANG Hui[1,2],LIU Xue-jie[1]

(1. *College of Computer Science and Technology, Jilin University, Changchun* 130012, *China*; 2. *Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education , Jilin University, Changchun* 130012, *China*)

Abstract: The limitations of current image manipulation localization methods, such as the extraction of features at a single scale, the misdetection and omission of small tampered regions caused by background confusion, and the uncertainty existing in prediction results, are addressed. An image manipulation localization method based on edge uncertainty learning is proposed. The base features of the tampered image are extracted by means of a pyramid vision transformer. A coarse localization map is then generated through multi-level interactive coarse localization branches. To enhance the detection of small tampered regions, a small target-aware refinement branch is employed. Multi-scale feature fusion is achieved with the use of a dedicated module, which enables the full interaction and integration of features across different scales. Additionally, entropy-based perceptual loss is introduced to supervise boundary uncertainty, thus

收稿日期:2025-01-06.

基金项目:国家重点研发计划项目(2022YFB4500600);吉林省科技发展计划重点研发项目(20230201088GX).

作者简介:陈海鹏(1978-),男,教授,博士.研究方向:图像处理与模式识别.E-mail:chenhp@jlu.edu.cn

通信作者:刘雪洁(1976-),女,副教授,博士.研究方向:信息安全.E-mail:xuejie@jlu.edu.cn

significantly reducing the uncertainty of the prediction results. The proposed method is evaluated on five widely-used public image tampering datasets in both in-domain and cross-domain experiments. It is demonstrated by the results that the method can effectively localize tampered regions and outperform existing approaches.

Key words: computer application; image manipulation localization; multi-scale feature interaction; refined perception of small tampering areas; boundary uncertainty learning

0 引 言

多媒体数字图像信息常被用于法律诉讼、刑事调查等多种重要场合[1]。随着图像处理技术的飞速发展,Photoshop、FakeApp等图像编辑工具被广泛普及[2]。不法分子趁机利用图像操纵手段从中获利,网络欺诈、虚假宣传、学术造假等现象层出不穷[3],造成了严重的负面影响。为更好地保障社会安全、维护网络秩序,开展图像篡改定位检测研究至关重要。

图像篡改主要包括拼接[4]、复制-粘贴[5]、移除[6]3种操作。早期的检测方法主要依赖手工提取特征,如彩色滤波阵列(Color filter array, CFA)[7]、JPEG压缩伪影[8]、噪声不一致性[9]等。上述方法在面对后处理攻击时,准确性和鲁棒性下降,无法应对当前飞速发展的图像处理技术。

近年来,深度学习技术的发展推动图像篡改定位检测任务不断向前。Chen等[10]通过显式监督的方式揭露篡改边界痕迹,以探索语义无关的篡改线索。Shi等[2]提出通过算子感应模块探索边界位置与篡改区域的互补关系,细化边界和区域特征。Shi等[11]提出利用双重注意力机制和边界监督网络定位篡改区域,以提高模型的检测性能和鲁棒性。尽管现有方法已取得显著成效,但仍存在一些问题:①特征尺度单一。仅从单一尺度学习和提取特征会导致上下文信息丢失,当面对不同尺度或尺度变化较大的篡改区域时,现有方法无法实现精准定位。②小篡改区域感知能力差。小篡改区域在图像中占据的像素较少,视觉特征变化细微,易导致前景与背景混淆。现有方法大多在固定尺度下进行局部特征学习,易出现区域误检、漏检情况。③预测结果存在不确定性。真实场景中的篡改图像较为复杂,现有方法大多采用二进制交叉熵损失与IoU损失组合训练,预测时会产生边界模糊和不确定性,降低模型可靠性。

为解决上述问题,本文提出了基于边界不确定性的渐进式边界不确定性学习网络(Progressive boundary uncertainty learning network, PBUL-Net),专注于多尺度特征的感知交互学习和边界不确定性学习,实现了对篡改区域的精准分割,尤其是小篡改区域。PBUL-Net的主要贡献如下:①设计多级交互粗定位分支,包含门控融合策略和多注意粗定位模块(Multi-attention coarse locating module, MCLM),在减少低级特征噪声干扰的同时,可获取粗定位特征,用于后续对多尺度特征交互融合模块进行调制。②提出小目标感知细化分支,先利用全注意力模块(Full attention module, FAM)对图像的全局特征建模,捕捉上下文中的细微线索,再利用特征感知增强模块(Feature aware enhancement module, FAEM)扩大感受野,增强局部特征,使模型更专注于篡改区域。③采用多尺度特征融合模块(Multi-scale feature fusion module, MFFM),充分交互和融合多尺度特征,捕捉不同层次信息。④提出基于熵的边界不确定性感知损失(Entropy-based uncertainty boundary aware loss, EUBAL),降低预测结果的不确定性,提高模型的可靠性。⑤本文在公开数据集上分别进行域内和跨域实验,定量和定性实验结果均表明,本文方法(PBUL-Net)优于现有方法。

1 本文方法

1.1 整体网络结构

本文方法的整体网络结构主要由4个部分组成:骨干网络、多级交互粗定位分支、小目标感知细化分支以及多尺度特征融合模块,具体结构如图1所示。首先,骨干网络采用PVTv2[12]提取4个多层次特征。其次,利用多级交互粗定位分支对篡改图像进行粗略定位,具体包含门控融合策略和多注意粗定位模块,可更全面地获取多种篡改线索。该分支获得的粗定位图后续会输入多尺

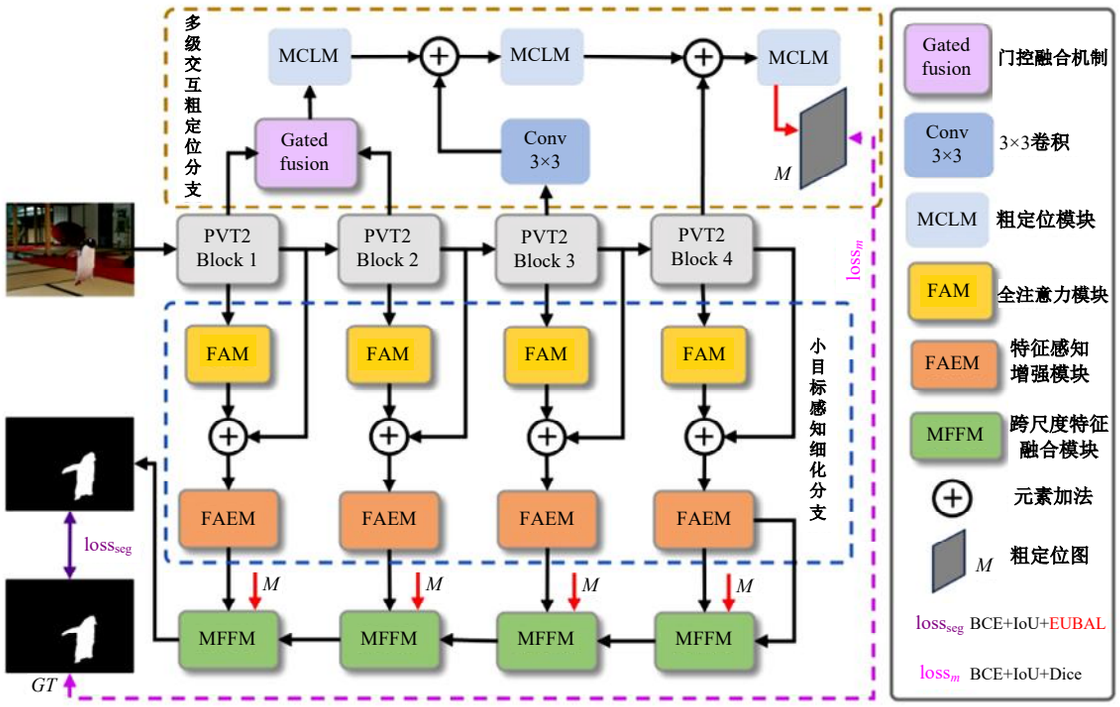

图 1　本文方法的整体网络结构

Fig. 1　Overall network structure of the proposed method

度特征融合模块进行调制,辅助模型精确预测。再次,利用小目标感知细化分支增强对小篡改区域的感知能力,其中全注意力模块用于增强同质性差异特征的表征能力,特征感知增强模块用于解决小篡改区域与背景易混淆的问题。最后,利用多尺度特征融合模块丰富多尺度特征表达。本文提出了基于熵的边界不确定性感知损失,以提高边界预测结果的可信度,并采用多级监督方法对模型进行训练。

1.2　多级交互粗定位分支

本文提出了利用多级交互粗定位分支对篡改图像进行粗略定位。该分支由门控融合策略和多注意粗定位模块组成。

1.2.1　门控融合策略

为降低低级特征中噪声对图像篡改任务的影响,本文提出利用门控融合策略抑制噪声,有效融合两层低级特征,并保留有用的纹理特征。首先,分别将两层低级特征输入 3×3 卷积层,经批归一化和激活函数处理后,得到两个新的特征图;其次,将两个特征映射串联并通过空间方向门控连接做进一步处理;最后,利用 Softmax 层生成权重,对特征进行加权融合。

1.2.2　多注意粗定位模块

为更全面地获取多种类型的线索,本文设计了多注意粗定位模块。该模块包含篡改注意分支、全局注意分支和局部注意分支 3 个组成部分,具体网络结构如图 2 所示。

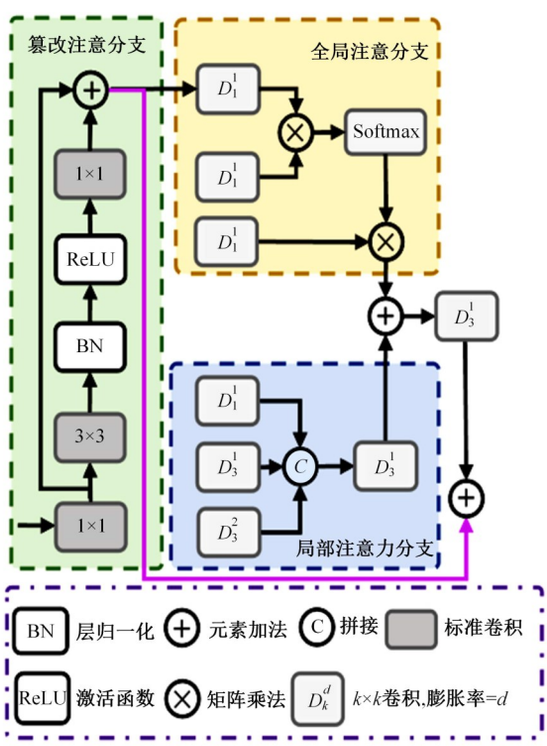

图 2　多注意粗定位模块结构

Fig. 2　Architecture of multi-attention coarse locating module

篡改注意分支通过逐步增强图像特征并进行有效的信息融合,帮助模型有效识别图像中潜在的篡改区域。通过1×1卷积层的通道转换和批归一化层的标准化处理去除冗余信息,避免了可能干扰检测的背景噪声和无关区域。

全局注意分支可帮助模型捕获丰富的上下文信息并增强语义表达。用3个1×1卷积层生成查询特征$\boldsymbol{Q}$、关键特征$\boldsymbol{K}$和值特征$\boldsymbol{V}$。假设输入特征映射为$\boldsymbol{F}\in\mathbb{R}^{c\times h\times w}$,其中$c$、$h$、$w$分别为通道数、宽度和高度。为节省计算开销,将$\boldsymbol{Q}$和$\boldsymbol{K}$的通道数设置为$c/4$,将$\boldsymbol{Q}$、$\boldsymbol{K}$重构为$\mathbb{R}^{(c/4)\times n}$,$\boldsymbol{V}$重构为$\mathbb{R}^{c\times n}$,其中$n=h\times w$。对$\boldsymbol{Q}$的转置和$\boldsymbol{K}$进行矩阵相乘,并用Softmax函数生成全局空间注意力图$\boldsymbol{GSA}_a\in\mathbb{R}^{n\times n}$。将$\boldsymbol{V}$和$\boldsymbol{GSA}_a$利用矩阵乘法相结合,得到语义增强特征$\boldsymbol{F}_e$。

局部注意分支使用多个卷积层挖掘局部细节。采用3个不同核大小的扩张卷积层,并通过串联聚合3层特征,随后将聚合特征输入3×3卷积层得到最终的局部输出$\boldsymbol{F}_l$。由于局部注意分支中的卷积层核尺寸相对较小,因此其可更多地关注边缘信息。

最后使用元素加法将全局语义增强特征$\boldsymbol{F}_e$与局部增强特征$\boldsymbol{F}_l$相加进行进一步细化,然后将二者连接,得到最终的输出$\boldsymbol{F}_{cl}$。对多注意粗定位模块进行多级交互,最终得到粗定位图$\boldsymbol{M}$,用于后续对多尺度特征融合模块进行调制。

1.3 小目标感知细化分支

为提高模型对小篡改区域的感知能力,本文提出小目标感知细化分支,以增强小目标的弱特征表征。该分支由全注意力模块和特征感知增强模块组成。

1.3.1 全注意力模块

图像篡改区域的形状是多种多样的,存在某些原始图像中部分对象被删除或背景被替换的情况。因此,本文提出全注意力模块,以增强同质性差异特征的表征能力,保持篡改区域的细节与语义一致性,具体结构如图3所示。

首先,使用大小为$H\times 1$和$1\times W$的池化核分别对输入特征进行全局平均池化和线性操作,得到两个不同维度的特征。其次,将两种特征进行裁剪和融合,得到中间特征。再次,将输入特征分别在H和W维度上分割成H个和W个切片,生成$\boldsymbol{K}$和$\boldsymbol{V}$,通过计算二者相似度得到注意力矩阵。最后,将注意力矩阵与$\boldsymbol{V}$相乘,使每个像素能感知水平和垂直方向上其他像素的信息,再乘以尺度参数γ,并加上原始输入特征得到最终增强后的特征。

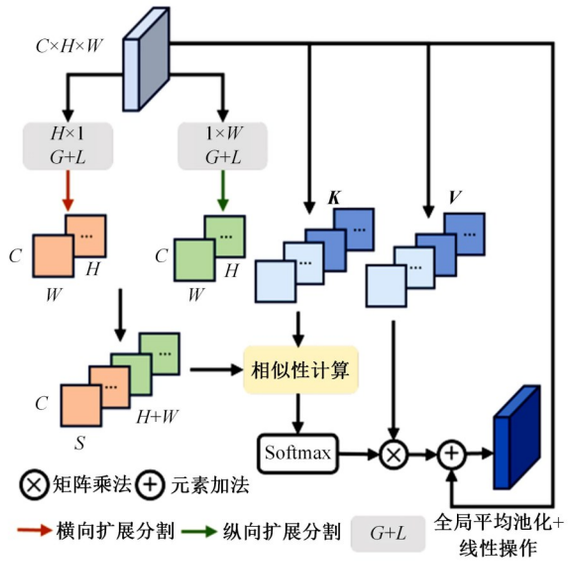

图3 全注意力模块结构

Fig. 3 Architecture of full attention module

1.3.2 特征感知增强模块

执行小篡改区域定位任务时,易发生篡改区域与背景混淆的伪预测问题。为此,本文提出特征感知增强模块,以增加特征的丰富性,并获取更丰富的局部上下文信息,具体结构如图4所示。

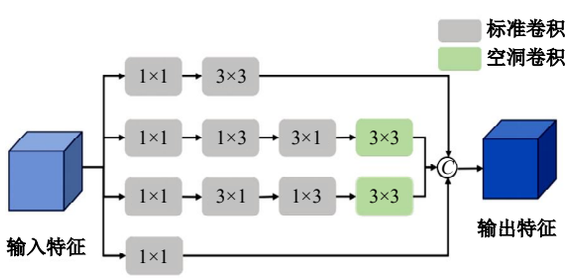

图4 特征感知增强模块结构

Fig. 4 Architecture of feature aware enhancement module

该模块设计了多分支卷积结构,使用标准卷积和扩张卷积获取丰富的局部上下文信息。每个分支对输入特征映射进行卷积,以调整通道数。第一个分支为残差结构,形成等效映射,可保留关键特征信息。其他3个分支通过不同尺寸的卷积核进行级联标准卷积操作,其核大小分别为1×3、3×1和3×3。其中,前两个分支额外添加了空洞卷积层,使提取的特征图能保留更多上下文信息。

1.4　多尺度特征融合模块

为丰富特征的尺度多样性表达,实现多尺度特征的充分交互,本文提出多尺度特征融合模块,具体结构如图5所示。

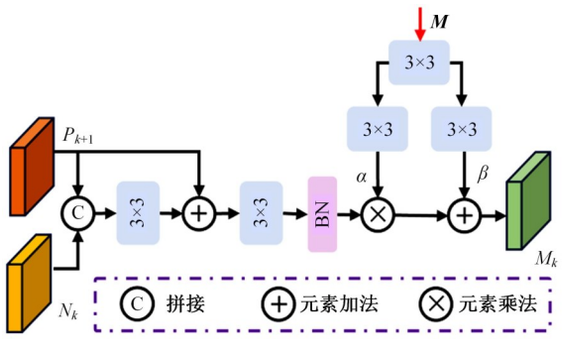

图5　多尺度特征融合模块结构

Fig. 5　Architecture of multi-scale feature fusion module

对于第k阶段($k\in\{1,2,3\}$),首先将模块下一阶段的输出P_{k+1}与当前阶段小目标感知细化分支获得的特征进行串联融合。其次,将融合结果输入3×3卷积层,通过添加P_{k+1}执行元素求和操作,再采用3×3卷积层进一步优化。最后,通过多级交互粗定位分支生成的粗定位图$\boldsymbol{M}$得到位置信息。模块采用条件归一化方法[13]将位置信息注入融合特征,将粗定位图$\boldsymbol{M}$输入3×3卷积层生成调制参数α和β。α和β通过卷积层动态生成,能根据输入特征自适应地调整特征的尺度和位置信息。为将位置信息更好地嵌入特征空间,该模块利用α和β作为仿射参数来调制特征,此过程可表示为:

$$M_k=\mathrm{BNC}(\mathrm{Cat}(P_{k+1},N_k)\oplus P_{k+1})\otimes \alpha(\boldsymbol{M})\oplus\beta(\boldsymbol{M}) \tag{1}$$

式中:BNC(•)为一个3×3批归一化卷积层;Cat(•)为串联操作;$k\in\{1,2,3\}$;$\otimes$为乘法运算;$\oplus$为加法运算。

当$k=4$时,将下一阶段模块对应的输入替换为第4个特征感知增强模块的输出。

1.5　基于熵的边界不确定性感知损失

当使用加权二元交叉熵(Binary cross-entropy,BCE)损失和加权IoU损失对篡改区域进行掩码监督时,模型预测会产生严重的模糊性和不确定性,无法准确定位篡改区域边界,从而降低方法的可靠性。因此,本文提出了基于熵的边界不确定性感知损失,增加对模糊预测的惩罚。

熵通常用于衡量分类任务中预测分布的不确定性。因此,本文将熵与不确定性感知损失(Uncertainty-aware loss,UAL)[14]相结合,提出基于熵的边界不确定性感知损失,利用预测值的熵衡量模型预测定位结果的不确定性,并对熵值较高的像素进行额外惩罚,此过程可表示为:

$$\mathcal{L}_{\mathrm{EUBAL}}=(1-\sigma)\times[-p_{i,j}\log(p_{i,j})-(1-p_{i,j})\cdot\log(1-p_{i,j})]+\sigma\times(1-|2p_{i,j}-1|^2) \tag{2}$$

式中:$\mathcal{L}_{\mathrm{EUBAL}}$为基于熵的边界不确定性感知损失;$\sigma$为平衡参数,本文采用余弦策略对参数进行调整;$p_{i,j}$为模型在$(i,j)$处预测篡改区域的概率,通过模型的Sigmoid函数输出直接得到,用于计算模型预测的不确定性。

总损失函数$\mathcal{L}_{\mathrm{total}}$可表示为:

$$\mathcal{L}_{\mathrm{total}}=\mathcal{L}_{\mathrm{BCE}}^{w}+\mathcal{L}_{\mathrm{IoU}}^{w}+\mathcal{L}_{\mathrm{EUBAL}} \tag{3}$$

式中:$\mathcal{L}_{\mathrm{BCE}}^{w}$、$\mathcal{L}_{\mathrm{IoU}}^{w}$分别为二元交叉熵损失和加权IoU损失。

此外,本文采用多级监督方式,对粗定位图采用加权二元交叉熵损失、加权IoU损失和Dice损失混合监督的方式,此过程可表示为:

$$\mathcal{L}_{M}=\mathcal{L}_{\mathrm{BCE}}^{w}+\mathcal{L}_{\mathrm{IoU}}^{w}+\lambda\mathcal{L}_{\mathrm{Dice}} \tag{4}$$

式中:$\mathcal{L}_{\mathrm{Dice}}$为Dice损失;$\lambda$为权衡参数,本文实验中设置$\lambda=3$。

2　实验结果及分析

2.1　实验数据集及评价指标

本文使用5个公开的标准图像篡改检测数据集验证模型性能,分别为CASIA[15]、NIST16[16]、COVER[17]、Columbia[18]、IMD2020[19]。为公平比较,本文进行了域内和跨域两种实验。域内实验对5个标准数据集采用标准的训练测试分割配置,这些数据集的详细情况如表1所示。跨域实验使用CASIA v2.0[15]作为训练集,分别在CASIA[15]、NIST16[16]和Columbia[18] 3个数据集上进行测试。

本文在像素级评估了本文图像篡改定位方法的性能,采用曲线下面积(Area under the curve,AUC)和F_1分数进行评估,数值越高,表示网络检测性能越好。

2.2　实验环境及参数设置

实验基于PyTorch深度学习框架,模型在单个NVIDIA GeForce RTX 3090 GPU上进行训练。图像输入尺寸统一调整为256×256。在训练阶段,批量大小设置为16,并采用Adam优化器

表1 数据集描述

Table 1 Description of the datasets

数据集	分割设置		篡改操作类型			后处理操作
	训练	测试	拼接	复制-粘贴	移除	
CASIA[15]	5 123	921	√	√		√
NIST16[16]	404	160	√	√	√	√
COVER[17]	75	25		√		√
Columbia[18]	130	50	√			
IMD2020[19]	1 610	400	√	√	√	√

对模型进行优化。初始学习率设定为2×10^{-5},并采用周期性衰减策略进行调整,每30个epoch以0.9的衰减率逐渐下降。

2.3 结果分析及对比

为验证本文方法执行图像篡改定位任务时的有效性和通用性,进行域内和跨域两种实验,并分析比较实验结果。

2.3.1 定量实验结果对比

为验证PBUL-Net执行图像篡改定位任务时的有效性,将PBUL-Net与最新方法ManTra-Net[20]、SPAN[21]、DenseFCN[22]、SATFL[23]、MVSS-Net[10]、TANet[2]、DMFF-Net[24]、DAE-Net[11]进行比较,域内实验结果如表2所示。由表2可以看出,相较于最新的DAE-Net[11],PBUL-Net的AUC平均提高2%,F_1平均提高12.7%,实现了大幅度的提升。这揭示了小目标感知细化与边界不确定性学习的优越性,同时也凸显出多尺度特征融合在提升模型对不同尺度物体感知精度中的重要性。

表2 本文方法与其他方法的域内实验定量结果比较

Table 2 Comparison of quantitative results of in-domain experiments between this paper's method and other methods

方法	CASIAV		NIST16		Columbia		COVER		IMD2020		Mean	
	AUC	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC	F_1	AUC	F_1
ManTra-Net[20]	0.648	0.223	0.795	0.462	0.824	—	0.777	0.283	0.785	0.265	0.766	0.308
SPAN[21]	0.709	0.213	0.961	0.582	0.936	0.815	0.791	0.325	—	—	0.849	0.484
DenseFCN[22]	0.631	0.209	0.954	0.704	0.881	0.710	0.754	0.185	0.723	0.286	0.789	0.419
SATFL[23]	0.697	0.246	0.937	0.613	0.892	0.804	0.767	0.347	0.796	0.300	0.818	0.462
MVSS-Net[10]	0.748	0.390	0.981	0.827	0.719	0.703	0.808	0.284	0.817	0.411	0.815	0.523
TANet[2]	0.739	0.493	0.955	0.901	0.982	0.960	0.756	0.425	0.766	0.383	0.840	0.633
DMFF-Net[24]	0.791	0.386	0.976	0.843	0.945	0.837	0.727	0.282	0.784	0.328	0.845	0.535
DAE-Net[11]	0.805	0.494	0.984	0.871	0.973	0.872	0.741	0.330	0.826	0.338	0.866	0.581
PBUL-Net	0.852	0.662	0.981	0.906	0.982	0.961	0.819	0.525	0.796	0.488	0.886	0.708

为探索PBUL-Net在未知数据集上的通用性,本文参照DAE-Net[11]的实验设置进行跨数据集评估实验。由表3可以看出:PBUL-Net在5个数据集上均展示出了最佳性能,AUC和F_1相较于最新的DAE-Net[11]均有大幅度提升,PBUL-Net可有效应对不同数据集中的多样性场景和复杂变化,具有广泛适用性,更适合在实际场景中应用。

表3 与其他方法在未知数据集上的定量结果比较

Table 3 Quantitative comparison of this paper's method with other method on unseen datasets

方法	CASIAV		NIST16		Columbia	
	AUC	F_1	AUC	F_1	AUC	F_1
ManTra-Net[21]	0.648	0.223	0.475	0.095	0.481	0.386
DenseFCN[22]	0.631	0.209	0.604	0.051	0.586	0.317
MVSS-Net[23]	0.748	0.390	0.643	0.246	0.697	0.471
TANet[2]	0.739	0.493	0.638	0.248	0.710	0.440
DAE-Net[12]	0.805	0.494	0.663	0.299	0.726	0.486
PBUL-Net	0.852	0.662	0.702	0.344	0.800	0.690

2.3.2 定性实验结果对比

篡改区域检测的可视化结果如图6所示,图中前3行篡改操作从上到下依次为拼接、复制-粘贴和移除,后3行为小篡改区域时的定位结果。从图6中可以看出,PBUL-Net对3种篡改操作均取能得很好的检测结果,充分展示了其优越性及通用性。从后3行定位结果可以看出,尽管小篡改区域前景与背景高度相似极易导致误检,但其仍可精准定位。此外,利用PBUL-Net可提高对边界的分割能力,这得益于边界不确定性学习机制。

2.3.3 消融实验

为验证各个模块的有效性,本文在CASIAV数据集上进行了消融实验。"a"代表基线网络Baseline,以PVTv2[12]作为骨干编码网络,解码网络采用掩码生成块,包括卷积、批归一化、ReLU激活函数和Sigmoid激活函数。在此基础上,逐级添

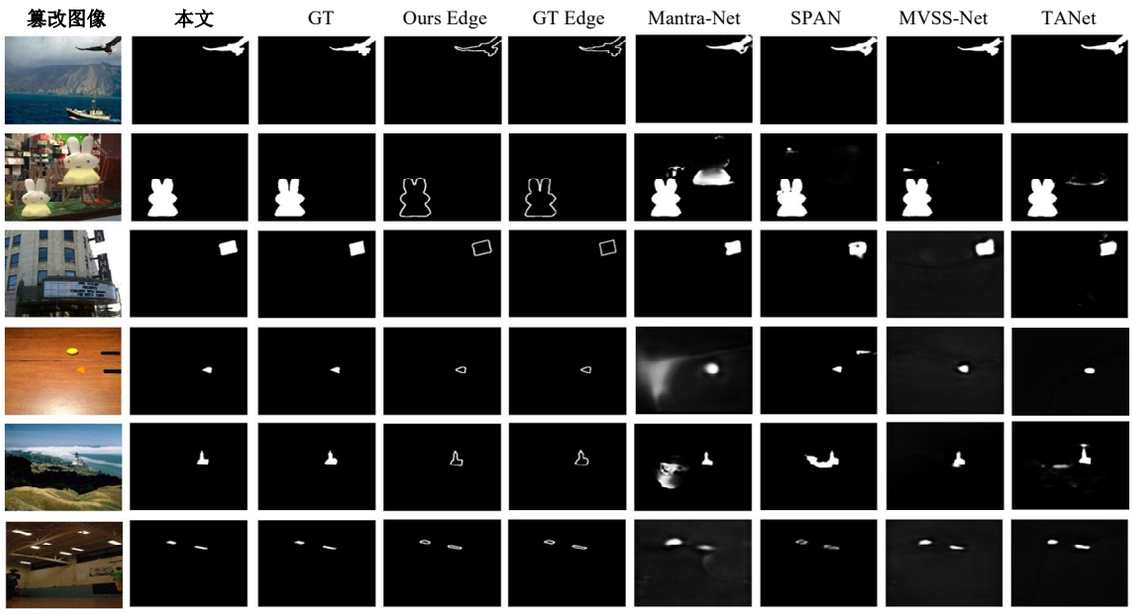

图 6　本文方法与其他方法的定性比较

Fig. 6　Qualitative comparison of this paper's method with other methods

加各模块以评估其有效性,实验结果如表 4 所示。

(1)多尺度特征融合模块有效性分析。由表 4 可知,与网络 a 相比,网络 b 引入多尺度特征融合模块后,AUC 由 0.810 提升至 0.819,F_1 分数由 0.605 提升至 0.618。这证明了模块使多尺度特征充分交互,提高了模型的表征能力。

(2)多级交互粗定位分支有效性分析。表 4 中 X 代表通过该分支得到的粗定位图。由表 4 可知,网络 c 在所有指标上均优于网络 b。这证明了利用粗定位图 X 对多尺度特征融合模块进行调制可有效提高模型的检测性能。

(3)全注意力模块有效性分析。通过比较网络 d 和网络 c 可以发现,引入全注意力模块后,AUC 提高了 0.006,F_1 分数提高了 0.012。这主要是因为,该模块增强了同质性差异特征的表征能力,从而提高了模型对篡改区域的检测能力。

表 4　不同组件在 CASIA 数据集的定量实验

Table 4　Quantitative results of different components on CASIA dataset

网络	组件					AUC	F_1
	MFFM	X	FAM	FAEM	EUBAL		
a						0.810	0.605
b	√					0.819	0.618
c	√	√				0.832	0.629
d	√	√	√			0.838	0.641
e	√	√	√	√		0.848	0.653
f	√	√	√	√	√	**0.852**	**0.662**

注:表中加粗字体表示该指标的最佳值。

(4)特征感知增强模块有效性分析。由表 4 可知,在特征感知增强模块的辅助下,网络 e 性能显著提升:AUC 由 0.838 提升至 0.848,F_1 分数由 0.641 提升至 0.653。这表明该模块可增强特征丰富性,有效解决前景与背景混淆造成的伪预测问题。

(5)基于熵的边界不确定性感知损失有效性分析。网络 f 的数据显示,F_1 分数由 0.653 提升至 0.662,AUC 由 0.848 提升至 0.852。这充分说明通过引入基于熵的边界不确定性感知损失能降低复杂背景干扰,成功减少模型对预测结果的不确定性,尤其是边界区域,实现了对篡改区域的精准定位。

2.4　鲁棒性分析

为验证本文方法的鲁棒性,对 NIST16 应用了 3 种图像失真方法,包括高斯模糊、JPEG 压缩以及高斯噪声,鲁棒性结果如图 7 所示。

从图 7 中可以看出,PBUL-Net 面对各种图像失真方法时均表现出最佳的鲁棒性。其中,ori 表示原始未受攻击时的模型性能。这主要是因为,模型有效利用多尺度特征的交互学习,融合了相邻层次特征,并针对易误检的小篡改区域进行感知细化。此外,引入基于熵的边界不确定性感知损失极大地降低了模型对预测结果的不确定性,提高了模型的抗干扰能力。

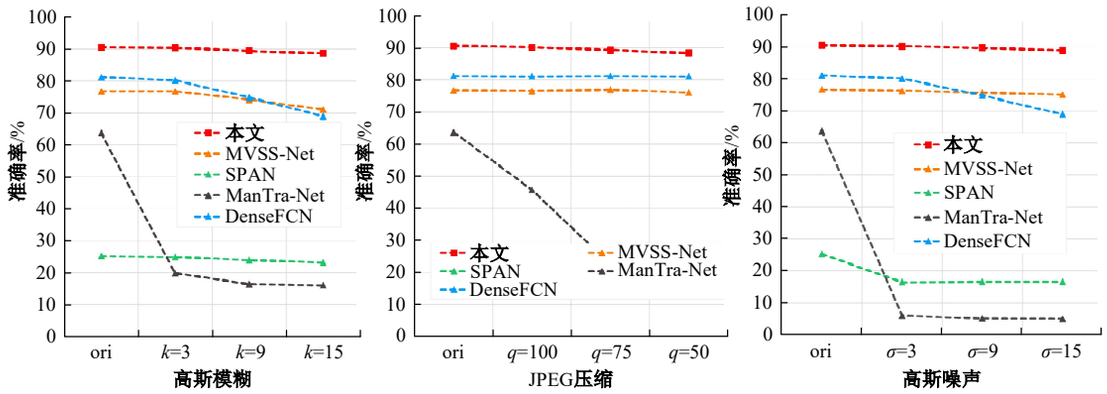

图 7 鲁棒性实验结果

Fig. 7 Robustness experiment results

3 结束语

本文针对现有图像篡改定位方法存在的问题,提出了基于边界不确定性学习的图像篡改定位方法,以及基于边界不确定性的渐进式边界不确定性学习网络,专注于多尺度特征的感知交互学习和边界不确定性学习,提高对小篡改区域的感知能力,并降低模型对预测结果的不确定性,实现了对篡改区域的精准分割。实验结果表明,本文方法在多个数据集上均显示出最佳性能,并在区分真实场景中各种原始图像与篡改图像方面具有卓越的泛化能力。未来将尝试对边界特征进行精细提取和整合,并将其作为先验指导,进一步提高模型对篡改区域定位的准确性和泛化性。

参考文献:

[1] 钟辉, 康恒, 吕颖达, 等. 基于注意力卷积神经网络的图像篡改定位算法[J]. 吉林大学学报: 工学版, 2021, 51(5): 1838-1844.
Zhong Hui, Kang Heng, Lyu Ying-da, et al. Image manipulation localization algorithm based on channel attention convolutional neural networks[J]. Journal of Jilin University (Engineering and Technology Edition), 2021, 51(5): 1838-1844.

[2] Shi Z, Chen H, Zhang D. Transformer-auxiliary neural networks for image manipulation localization by operator inductions[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(9): 4907-4920.

[3] 石泽男, 陈海鹏, 张冬, 等. 预训练驱动的多模态边界感知视觉 Transformer[J]. 软件学报, 2023, 34(5): 2051-2067.
Shi Ze-nan, Chen Hai-peng, Zhang Dong, et al. Pretraining-driven multimodal boundary-aware vision transformer[J]. Journal of Software, 2023, 34(5): 2051-2067.

[4] Liu Y, Zhu X, Zhao X, et al. Adversarial learning for constrained image splicing detection and localization based on atrous convolution[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(10): 2551-2566.

[5] Chen B, Tan W, Coatrieux G, et al. A serial image copy-move forgery localization scheme with source/target distinguishment[J]. IEEE Transactions on Multimedia, 2020, 23: 3506-3517.

[6] Zhang Y, Fu Z, Qi S, et al. Localization of inpainting forgery with feature enhancement network[J]. IEEE Transactions on Big Data, 2022, 9(3): 936-948.

[7] Popescu A C, Farid H. Exposing digital forgeries in color filter array interpolated images[J]. IEEE Transactions on Signal Processing, 2005, 53(10): 3948-3959.

[8] Bianchi T, Piva A. Image forgery localization via block-grained analysis of JPEG artifacts[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 1003-1017.

[9] Mahdian B, Saic S. Using noise inconsistencies for blind image forensics[J]. Image and Vision Computing, 2009, 27(10): 1497-1503.

[10] Chen X, Dong C, Ji J, et al. Image manipulation detection by multi-view multi-scale supervision[C]// Proceedings of the IEEE/CVF. International Conference on Computer Vision. Montreal, QC, Canada, 2021: 14165-14173.

[11] Shi C, Wang C, Zhou X, et al. DAE-net: Dual attention mechanism and edge supervision network for image manipulation detection and localization[J]. IEEE Transactions on Instrumentation and Measure-

ment, 2024(73): No. 5028112.

[12] Wang W, Xie E, Li X, et al. Pvt v2: Improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415-424.

[13] Park T, Liu M Y, Wang T C, et al. Semantic image synthesis with spatially-adaptive normalization[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA: USA, 2019: 2337-2346.

[14] Pang Y, Zhao X, Xiang T Z, et al. Zoom in and out: A mixed-scale triplet network for camouflaged object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA: USA, 2022: 2160-2170.

[15] Dong J, Wang W, Tan T. Casia image tampering detection evaluation database[C]//2013 IEEE. China Summit and International Conference on Signal and Information Processing. Beijing: China, 2013: 422-426.

[16] Guan H, Kozak M, Robertson E, et al. MFC datasets: Large-scale benchmark datasets for media for ensic challenge evaluation[C]//2019 IEEE. Winter Applications of Computer Vision Workshops (WACVW). Waikoloa, HI: USA, 2019: 63-72.

[17] Wen B, 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, 2016: 161-165.

[18] Hsu Y F, Chang S F. Detecting image splicing usinggeometry invariants and camera characteristics consistency[C]//2006 IEEE. International Conference on Multimedia and Expo. Toronto, ON: Canada, 2006: 549-552.

[19] Novozamsky A, Mahdian B, Saic S. IMD2020 a largescale annotated dataset tailored for detecting manipulated images[C]//Proceedings of the IEEE/CVF. Winter Conference on Applications of Computer Vision Workshops. Snowmass Village, CO: USA, 2020: 71-80.

[20] 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, 2019: 9543-9552.

[21] Hu X, Zhang Z, Jiang Z, et al. SPAN spatial pyramid attention network for image manipulation localization[C]//Proceedings of the European Conference on Computer Vision (ECCV). Glasgow: UK, 2020: 312-328.

[22] Zhuang P, Li H, Tan S, et al. Image tampering localization using a dense fully convolutional network [J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2986-2999.

[23] Zhuo L, Tan S, 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.

[24] Xia X, Su L C, Wang S P, et al. DMFF-net: Double-stream multilevel feature fusion network for image forgery localization[J]. Engineering Applications of Artificial Intelligence, 2024, 127: No. 107200.