DOI:10.19651/j.cnki.emt.2520043

基于多尺度跨层融合的图像篡改定位*

郭仕佳 张宝林 马琬雲 张雨初
(河北工业大学电子信息工程学院 天津 300401)

摘 要: 恶意篡改图像给人们的生活和社会带来严重影响。目前有许多模型用于图像篡改检测,但是存在细节信息丢失、小目标检测能力不足等问题。对于以上问题,提出了一种基于多尺度跨层融合的图像篡改定位模型。该模型在 RRU-Net 基础上改变了跳跃连接方式,设计 MSC-SC 结构,通过空洞空间金字塔池化强化跳跃连接中的编码器特征,跨层融合模块自适应融合编码器特征图与解码器特征图;另外,引入三重注意力机制增强下采样过程中的特征感知能力;最后,采用联合损失函数缓解正负样本不平衡问题。在 CASIA v2 和 COLUMB 数据集上的实验表明,所提方法相较于原始 RRU-Net,$F1$ 分数分别提升 11.33%和 6.84%,表明本检测方法的效果显著。

关键词: 图像篡改检测;RRU-Net;跳跃连接;三重注意力;损失函数
中图分类号: TP391.4;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.40

Image tampering localization based on multi-scale cross-layer fusion

Guo Shijia Zhang Baolin Ma Wanyun Zhang Yuchu
(College of Electronic Information Engineering, Hebei University of Technology, Tianjin 300401, China)

Abstract: Maliciously tampered images pose serious threats to both daily life and society. Although numerous detection models have been developed for image tampering detection, they still suffer from issues such as the loss of fine details and insufficient capability in detecting small targets. To address these challenges, this study proposes a multi-scale cross-layer fusion-based for image tampering localization. Built upon the RRU-Net framework, the model changes the skip connection scheme and enhances the encoder features in skip connections using the MSC-SC structure with atrous spatial pyramid pooling. A cross-layer fusion module is then designed to adaptively fuse the encoder and decoder feature maps. In addition, a triple attention mechanism is introduced to enhance feature perception during the down sampling process. Finally, a joint loss function is utilized to alleviate the imbalance between positive and negative samples. Experiments conducted on the CASIA v2 and COLUMB datasets demonstrate that the proposed method achieves $F1$ score improvements of 11.33% and 6.84%, respectively, compared with the original RRU-Net, indicating that significant effectiveness is achieved.

Keywords: image tampering detection;RRU-Net;skip connection;triplet attention;loss function

0 引 言

数字图像技术的发展拓宽了信息传播的渠道,依靠其直观性、丰富性、易于传播等特点,成为当今社会不可或缺的信息载体。随着数字图像编辑工具的发展,篡改图像的生成变得愈发容易,这对新闻真实性、司法取证、金融凭证审核等关键领域构成了严峻挑战[1]。

图像的篡改类型一般分为 3 类:拼接、复制粘贴和内容填充[2]。拼接指的是将一张图片的某个部分拼接到另外一张图片中;复制粘贴指的是一张图片内的某个要素移动复制到另外的区域;内容填充指的是在图像中填充和整体相匹配的内容。

传统方式的篡改图像检测技术主要依赖于手工特征提取和一些经典的图像处理方法来识别图像中的异常或篡改痕迹[3]。杨超等[4]提出一种基于篡改区域轮廓的拼接图像检测方法,通过检测异常边缘区域得到篡改区域,属于机器学习检测方法。近些年,众多研究者关于深度学习的图像篡改检测提出了许多办法,其本质是利用深度学习方法实现图像处理,目前基于深度学习图像检测与处理发展迅速,不少学者提供了许多的研究方法[5-6]。2016 年 Rao 等[7]将

收稿日期:2025-10-15
* 基金项目:河北省自然科学基金(F2023202013)项目资助

卷积神经网络(convolutional neural network,CNN)引入到图像篡改检测中,能够识别篡改区域但是无法对篡改区域定位。吴鹏等[8]提出一种改进的双流 Faster R-CNN 的检测定位算法,联合全卷积网络分支和错误等级分析算法实现篡改区域像素级定位,比原本算法精度更高,但仍有改进空间。周兴超等[9]提出一种边缘增强(edge feature enhancement module,EFEM)和双注意力机制的图像检测模型,但是没有针对小目标和细节部分专门建模,对于复杂背景的检测能力不足。张雷等[10]提出一种多尺度内容感知的图像篡改定位方法,通过多尺度内容感知融合和深度可分离卷积解码器实现图像篡改区域定位,但是该方法的检测结果存在边缘虚化、模糊等情况,边缘检测精度不够。喻小芹等[11]提出一种亮度对比度扰动下的图像篡改网络,能够抗击光照和对比度带来的干扰,在亮度扰动测试集有显著地提升,但是该模型的检测对象相对狭窄,只针对亮度和对比度处理后的篡改图片,并且不同的扰动因子下存在不同程度的图像误判情况。Chen 等[12]提出一种极简网络,只提取低层特征,避免高层语义信息的干扰,在复制粘贴数据集上有不错的表现,但是由于缺乏高级语义,使得网络在大面积篡改的数据集上表现稍逊。Sharma 等[13]提出一种基于特征融合和混合群智能优化的图像篡改检测方法,通过结合多目标粒子群优化(multi-objective particle swarm optimization,MOPSO)和海洋捕食者算法(marine predators algorithm,MPA)两种优化策略提高了模型的检测精度和鲁棒性,但是该方法只能对图像二分类,无法端到端的实现篡改区域的定位。Bi 等[14]在 U-Net[15]的网络基础上提出 RRU-Net 用于拼接图像的篡改检测,实现了端到端的篡改图像检测和像素级的篡改图像定位,但是简单的跳跃连接和连续的下采样导致特征图被压缩,对细节、小目标等区域感知能力下降,导致小目标篡改区域存在误判和识别不充分。

针对以上问题,本文提出一种基于多尺度跨层融合的图像篡改检测模型,用于拼接类型的图像检测。本文主要工作包括:

结合多尺度的跨层跳跃连接(multi-scale cross-layer skip connections,MSC-SC)。引入空洞空间金字塔池化(atrous spatial pyramid pooling,ASPP)[16],通过 3 个小膨胀率的空洞卷积提取编码器特征,增强编码器的细节保持能力,设计跨层融合模块(cross-level fusion,CLF)动态融合编码器和解码器特征图,提高模型对细节和小目标的检测能力。

三重注意力机制(triplet attention,TA)[17]。在下采样过程中引入轻量级注意力机制,通过三分支结构捕获通道与空间维度的跨维度交互,显著提升模型对篡改像素的感知能力。

联合损失函数(dice-BCE loss,DBL)。采用二元交叉熵损失函数(binary cross entropy,BCE)与 Dice 联合损失函数,缓解样本不平衡问题,提升模型鲁棒性和稳定性。

1 RRU-Net 网络

RRU-Net 沿用了 U-Net 的基本结构,由编码器、解码器和跳跃连接组成,该网络创新性的提出了环形残差结构,在编码器和解码器中都添加了一个残差反馈,如图 1 所示。残差反馈利用未篡改与篡改区域之间的本质特征差异,提升篡改区域的可区分性,提升了网络的表征能力。除此之外 RRU-Net 还通过空洞卷积(dilated convolution,DC)和环形残差机制替代普通卷积模块,增强了特征表达能力和信息还原能力。

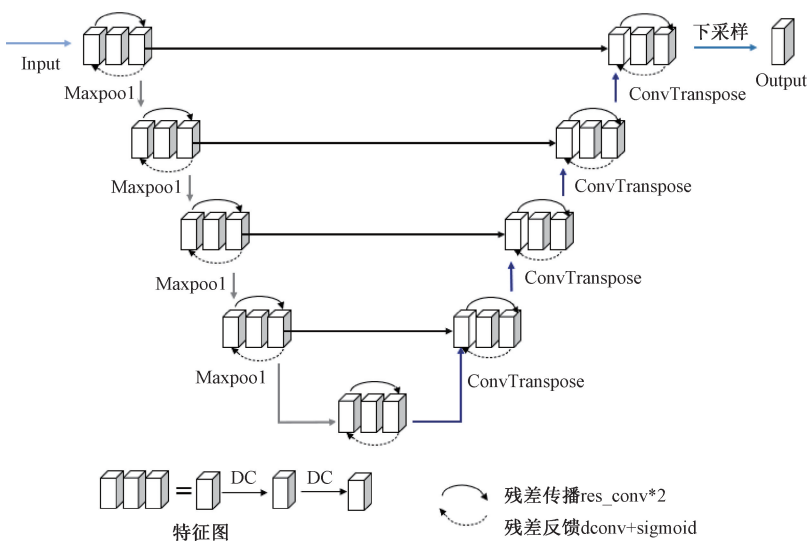

图 1 RRU-Net 模型结构

Fig. 1 Structure of the RRU-Net model

2　模型结构

2.1　框架概述

本文提出的基于多尺度跨层融合的图像篡改定位模型结构如图 2 所示。本模型以 RRU-Net 为基本框架，首先改变了跳跃连接方法，提出了 MSC-SC 结构，利用 ASPP 方法提取多尺度特征，提高了编码器的细节表达能力，并且设计了 CLF 模块用于层级融合，能够让网络自适应学习各个特征图的重要程度，以此达到动态融合；其次，为了增强模型的小目标检测效果，在编码器下采样过程中添加了 TA 模块，以提高模型的细节感知能力和检测精度。

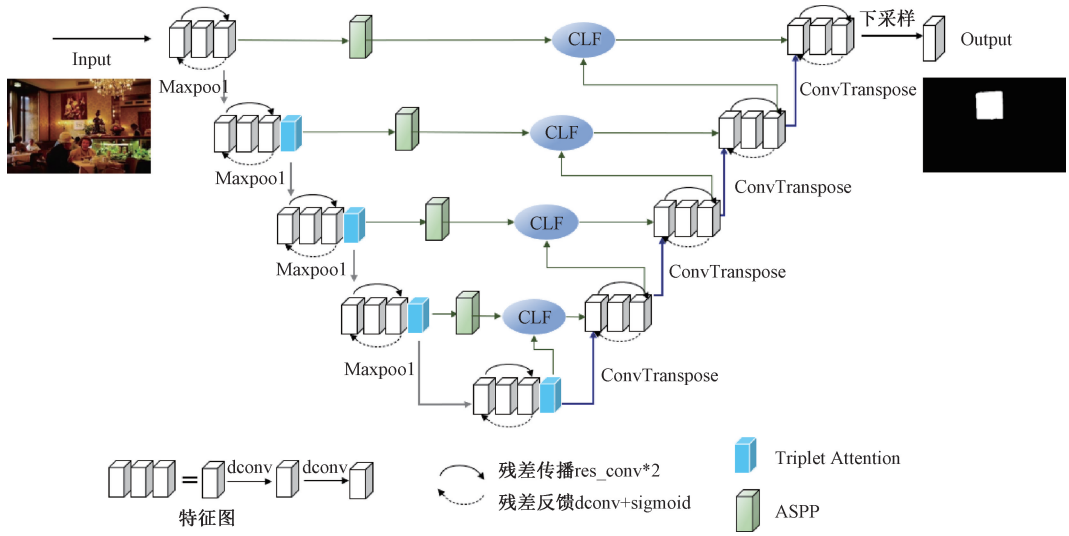

图 2　基于多尺度跨层融合的图像篡改定位模型结构

Fig. 2　Structure of the image tampering localization model based on multi-scale cross-layer fusion

2.2　MSC-SC

跳跃连接结构用于缓解梯度消失，构建更深的网络。为了将信息传递到更深层，Huang 等[18]提出了密集连接的卷积网络，将每一层进行短连接，使每一层都能直接接受所有前面层的特征。不过这种将所有层连接起来的方法相对激进，本文借助空洞空间金字塔池化提出一种 MSC-SC 结构，仅在原本的跳跃连接基础上连接更深层的解码器特征，用于增强模型对细节的检测效果，其整体结构如图 3 所示。

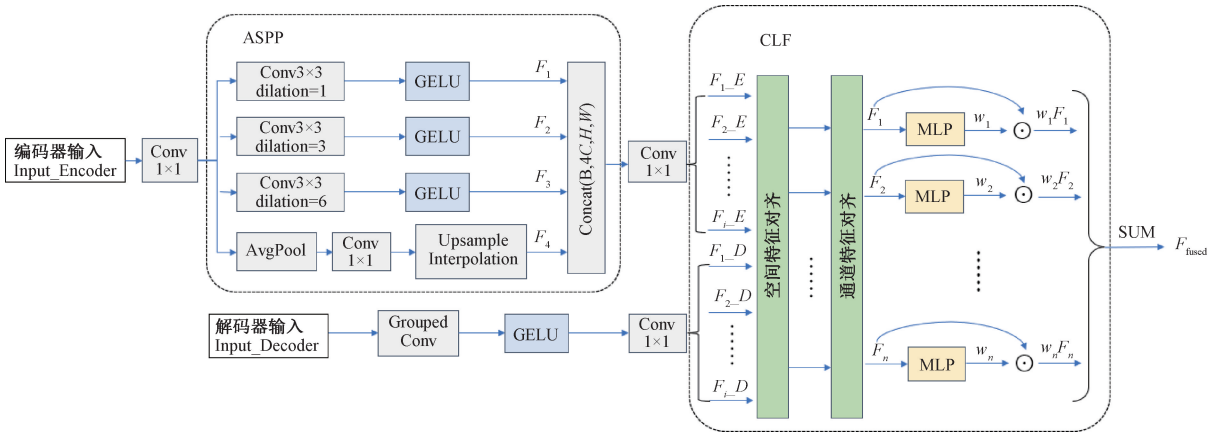

图 3　MSC-SC 结构

Fig. 3　MSC-SC module

不同层的编码器和解码器在传输中包含的信息并不相同，编码器中包含更多的细节和纹理特征，而深层的解码器则包含更多的语义信息。为了提高模型对细节的检测精度，提高模型的检测性能，使用深度可分离卷积对解码器特征进行特征增强，其中的分组卷积能够在空间尺寸不变的情况下提取局部信息，而编码器信息则借助 ASPP 增强特征图中的细节信息。

ASPP 包含空洞卷积和池化金字塔，通过不同膨胀率的空洞卷积获取不同感受野的特征。它拥有两种分支：一种是空洞卷积，通过不同的膨胀率提取多尺度信息；一种

是全局池化,用于捕获全局上下文语义。其计算公式为:

$$\boldsymbol{F}_k = \mathrm{GELU}(\boldsymbol{W}_{k*d}\boldsymbol{X}) \tag{1}$$

式中:$\boldsymbol{X} \in \mathbf{R}^{B\times C\times H\times W}$;$B$ 是批大小;C 是通道数;H 和 W 是高和宽;$*d$ 是膨胀率;$\boldsymbol{W}_k$ 是卷积核。

空洞卷积能够在不丢失分辨率的情况下扩大感受野,实现精准目标地定位。标准的 ASPP 拥有 5 条分支,分别为:1 条 1×1 的卷积、3 条空洞卷积分支和 1 条全局池化分支,空洞卷积的膨胀率分别为 6、12、18。为了小目标的检测效果,本文使用膨胀率为 1 的分支替代 1×1 的卷积,其余分支膨胀率设置为 3 和 6。

由于编码器和解码器特征包含的语义信息并不相同,为了让网络能够关注重要信息,抑制无关信息,需要一个灵活的融合方法。Li 等[19]提出一种动态选择性内核(selective kernel convolution, SK),使神经元能够根据输入内容自适应调整感受野大小。受此启发,本文提出一种 CLF 模块,用于融合编码器特征图和解码器特征图。

编码器和解码器传输来的特征张量在通道数上和尺寸大小上并不相同,所以需要通道对齐和插值上采样的方式尺寸对齐。设传入 CLF 的编码器特征为[$\boldsymbol{F}_{1\_}\boldsymbol{E}$, $\boldsymbol{F}_{2\_}\boldsymbol{E}$,…, $\boldsymbol{F}_{i\_}\boldsymbol{E}$],解码器特征为[$\boldsymbol{F}_{1\_}\boldsymbol{D}$, $\boldsymbol{F}_{2\_}\boldsymbol{D}$,…, $\boldsymbol{F}_{i\_}\boldsymbol{D}$],通过空间特征对齐和通道特征对齐得到张量和通道数一致的特征图[$\boldsymbol{F}_1$, $\boldsymbol{F}_2$,…, $\boldsymbol{F}_n$]。MLP 结构能够通过特征变换、非线性映射、信息融合和特征选择的过程筛选和增强关键区域特征,特征图信息经过 MLP 处理得到权重向量[$\boldsymbol{w}_1$, $\boldsymbol{w}_2$, $\boldsymbol{w}_3$,…, $\boldsymbol{w}_n$],最后通过加权融合得到最终特征图,计算公式为:

$$\boldsymbol{F}_{\mathrm{fused}} = \sum_{i=1}^{N} \boldsymbol{w}_i \boldsymbol{F}_i \tag{2}$$

SK 关注不同卷积核尺度提取的特性,而本文基于不同层级提取特性。通过不同层级特征自适应加权融合,能够让网络自适应的根据输入特征的内容来决定各层级特征的融合比例,增强模型对细节的敏感度,同时提高模型的泛化能力。

2.3 TA 注意力机制

为了进一步增强模型的细节提取能力,突出编码器的细节特征,本文在下采样过程中添加了 TA 注意力机制,位置如图 4 所示。

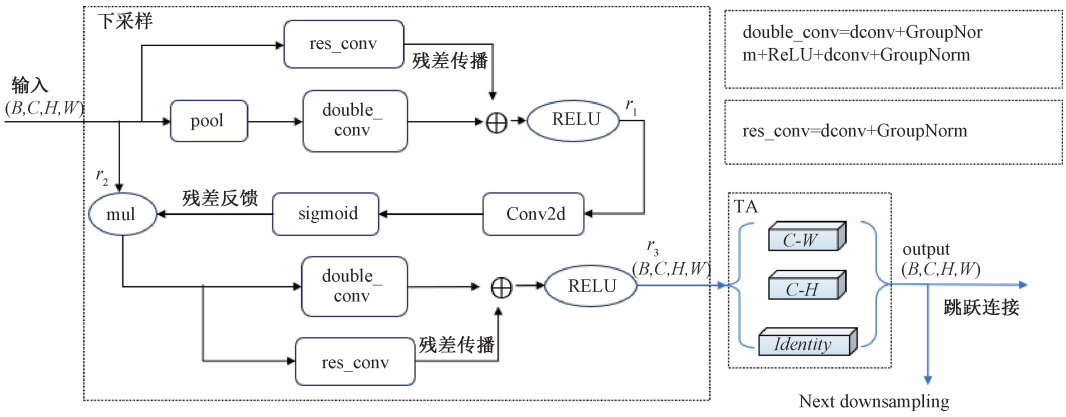

图 4 下采样模块

Fig. 4 Downsampling module

由于沿用了 RRU-Net 的基本结构,下采样中存在一个环形残差结构,其中包括两次残差传播和一次残差反馈。TA 模块的位置位于整个下采样之后,能够通过三分支结构同时关注通道与空间特征,实现协同感知,比如今主流的 CBAM(convolutional block attention module, CBAM)[20]注意力机制更轻量且高效,其内部结构如图 5 所示。

TA 模块通过 3 个并行分支分别处理不同维度的交互,顶层计算通道(C)和空间宽度(W)的注意力权重;中层处理通道(C)和空间高度(H)的注意力权重;底层用于捕获空间相关性(H-W)。在建立维度关系时,将输入张量顺时针或逆时针旋转,使得对应维度相互对齐,通过 Z-Pool 压缩后计算注意力的权重,最终 3 个通道取平均值。

以 C-W 分支为例,输入张量为 $\boldsymbol{\chi}_{c-w}=(B,C,H,W)$,通过维度旋转得到($B$,$H$,$C$,$W$),此时一条分支以 H 维度经过 Z-Pool($\boldsymbol{\chi}_{c-w}$)压缩处理,Z-Pool($\boldsymbol{\chi}_{c-w}$)的计算方法如式(3)所示。

$$\text{Z-Pool}(\boldsymbol{\chi}_{c-w}) = [\mathrm{MaxPool}(\boldsymbol{\chi}_{c-w}), \mathrm{AvgPool}(\boldsymbol{\chi}_{c-w})] \tag{3}$$

式中:MaxPool 表示最大池化;AvgPool 表示平均池化;$\boldsymbol{\chi}_{c-w}$ 的张量维度为(B,H,C,W)。

处理后的信息通过卷积生成注意力后由 Sigmoid 获得权重,与初始的(B,H,C,W)相乘获得注意力加权后的信息,最后再经过一次旋转将(B,H,C,W)旋转为(B,C,H,W),使输入输出的张量维度不变。

TA 模块通过 3 个分支分别捕获不同维度的交互,避免单一注意力机制导致的特征过度平滑。Z-Pool 操作在保留关键信息的同时压缩冗余特征,旋转对齐则确保跨维

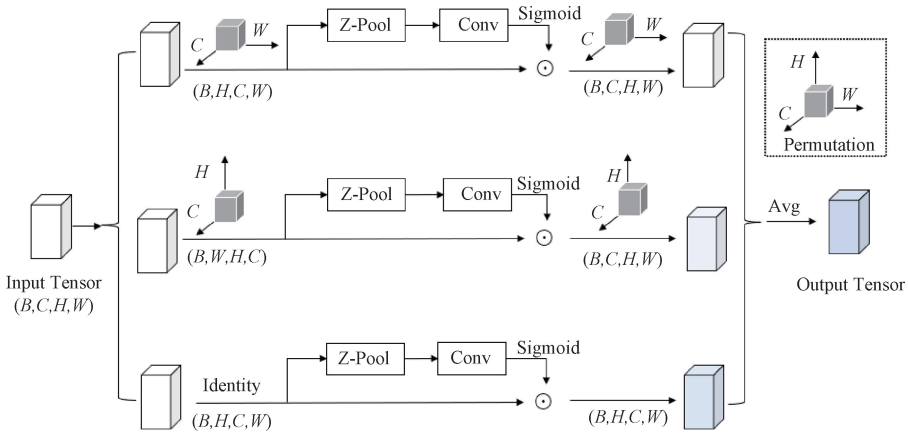

图 5　TA 模块结构

Fig. 5　Structure of the TA module

度交互时不丢失空间结构信息，从而提升模型对细节特征的感知能力。

2.4　联合损失函数

在深度学习训练模型中，损失函数在模型训练中起关键作用，用于评估预测结果与真实标签之间的误差，其值越小表明模型性能越优。在训练过程中，根据得到的损失反向传播计算梯度，优化器据此更新参数，从而指导模型参数的更新方向。

在二分类任务中，通常采用 BCE 作为损失函数，定义为：

$$L_{\mathrm{BCE}}=-\frac{1}{N}\sum_{i=1}^{N}[y_i\log(p_i)+(1-y_i)\log(1-p_i)] \tag{4}$$

式中：y_i 表示真实标签，该值由掩码图像提供，图像中的白色为正样本，用 1 表示，黑色是负样本，用 0 表示，p_i 表示模型预测的概率，N 表示像素总数。

BCE 损失函数的优势在于每个像素都独立评估，简单易优化，但是对于样本不平衡的场景，例如背景像素占比远高于目标像素，模型容易出现偏差，倾向于将大部分像素判定为背景像素，导致模型的泛化能力较差。

为解决以上问题，引入 Dice 损失函数。Dice 损失函数的定义为：

$$L_{\mathrm{Dice}}=1-\frac{2\sum_{i=1}^{N}y_ip_i}{\sum_{i=1}^{N}y_i+\sum_{i=1}^{N}p_i} \tag{5}$$

式中：y_i 和 p_i 表示真实标签和模型预测的概率，N 表示像素总数。

Dice 损失函数直接优化目标区域的重叠面积，不受背景像素数量的影响，即使目标像素很少，Dice 损失依然能强制模型关注其存在，常用于前景和背景不平衡的分割任务，对于图像篡改检测任务也十分受用。但是由于 Dice 损失优化的是区域重叠而非分类概率，可能输出非概率值，并且输出稳定性较差，在训练初期可能会出现预测全 0 或 1 的情况。

鉴于以上损失函数的特点，本文提出一种 DBL 损失函数：

$$L_{\mathrm{DBL}}=\alpha L_{\mathrm{Dice}}+\beta L_{\mathrm{BCE}} \tag{6}$$

式中：L_{Dice} 为 Dice 损失函数，L_{BCE} 为 BCE 损失函数。

联合损失函数既可以关注像素级的分类精度，又可以优化整体区域的匹配程度，而 BCE 的平滑梯度弥补 Dice 的梯度突变问题，能够在类别不平衡和结构约束任务中均表现良好。经过多次实验表明，α 取值为 0.2，β 取值为 0.8，模型的梯度最为稳定，效果最佳。

3　实验结果及分析

3.1　数据集及实验设置

本文对 CASIA v2、COLUMB 两个公开数据集进行训练和测试。CASIA v2 是由中国科学院自动化研究所发布的图像篡改检测数据集，是目前图像篡改检测使用的主流数据集之一，其中既包含拼接篡改类型，也包含复制粘贴篡改类型，为了研究拼接篡改类型的检测效果，从中选取 1 292 张拼接图片（tampered picture，TP）和 1 292 张对应的掩码图片（Mask）用于实验。

COLUMB 数据集是哥伦比亚大学创建的数据集，图像来源于数码相机拍摄，篡改区域相比于 CASIA v2 更大更为明显，只包含拼接篡改方式，不包含复制粘贴篡改方式。

为了增强模型的泛化性，使用水平翻转和垂直翻转的方法进行数据增强，数据集的数量扩大为原来的 3 倍，扩大后的拼接图片和掩码图片各有 3 876 张，按照 7∶2∶1的比例将数据集划分为训练集、验证集和测试集，具体数值如表 1 所示。并且在实验中将数据集中的图片调整为统一的大小，其中 CASIA v2 的图片大小处理为 384×256，COLUMB 的图片大小处理为 757×568。

本实验由 Python 3.8 作为编程语言，实验平台为 Ubuntu

表 1 实验数据集

Table 1 Experimental datasets 张

数据集	训练集		验证集		测试集	
	TP	Mask	TP	Mask	TP	Mask
CASIA v2	2 713	2 713	755	755	408	408
COLUMB	366	366	104	104	52	52

系统,显卡使用 NVIDIA GeForce RTX 4090D,CUDA 版本 11.1。实验优化器采用 Adam,初始学习率为 0.001,批处理大小为 16,共训练 100 个轮次。

3.2 实验评估

对于图像拼接篡改检测任务,像素级别的预测正误和篡改区域的准确定位是较为关注的问题。为评估所提出的检测方法,使用精确率(precision)、召回率(recall)和 $F1$ 进行性能评估,计算公式为:

$$\text{Precision}=\frac{\text{TP}+\text{epsilon}}{\text{TP}+\text{FP}+\text{epsilon}} \tag{7}$$

$$\text{Recall}=\frac{\text{TP}+\text{epsilon}}{\text{TP}+\text{FN}+\text{epsilon}} \tag{8}$$

$$F1=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}+\text{epsilon}} \tag{9}$$

式中:TP 为预测值为 1,且真实标签也为 1 的情况;FP 为预测值为 1,但真实标签为 0 的情况;FN 为预测值为 0,但真实标签为 1 的情况,epsilon 是一个常数,值为 1×10^{-7},目的是防止出现除零错误。

Precision 衡量模型的准确性,即预测为篡改的区域中,有多少是真正的篡改;Recall 衡量模型的查全率,即所有真实篡改区域中,有多少被成功检测出来;$F1$ 是评价模型性能的综合指标,兼顾准确率和召回率两个参数,最大为 1,最小为 0,值越大代表模型的性能越好。

3.3 消融实验

为验证联合损失函数、融合多尺度的跨层跳跃连接、三重注意力机制的有效性,在 CASIA v2 和 COLUMB 分别进行消融实验,通过逐一增加这些模块,直观体现这些模块各自性能的提升,实验结果如表 2 所示。

表 2 消融实验结果

Table 2 Results of ablation experiments

序号	MSC-SC	TA	DBL	CASIA v2			COLUMB		
				Precision	Recall	$F1$	Precision	Recall	$F1$
1				0.868 4	0.721 9	0.788 4	0.946 1	0.894 2	0.919 4
2	√			0.885 1	0.782 6	0.830 7	0.962 3	0.897 9	0.929 0
3		√		0.860 1	0.798 3	0.828 1	0.951 2	0.928 4	0.939 6
4			√	0.816 2	0.852 2	0.833 8	0.947 4	0.983 6	0.965 2
5	√		√	0.895 7	0.902 8	0.899 2	0.980 2	0.986 6	0.983 4
6		√	√	0.882 9	0.861 5	0.872 1	0.968 7	0.974 2	0.971 4
7	√	√		0.902 6	0.865 3	0.883 5	0.978 5	0.974 1	0.976 3
8	√	√	√	0.912 7	0.891 1	0.901 7	0.989 6	0.987 9	0.987 8

通过表 2 的消融实验,能够看到每个模块对性能的提升。首先看到改变了损失函数的定义,使用联合损失函数,Recall 有了明显的提升,其次通过 MSC-SC 模块和 TA 模块,模型的 Precision 和 Recall 都有了显著的提升。

总体来说,所提方法在 CASIA v2 和 COLUMB 数据集上相较于原始 RRU-Net,$F1$ 分数分别提升 11.33% 和 6.84%。

3.4 对比实验

为了证明改进算法的性能,选取了几种经典的语义分割模型 U-Net、RRU-Net、DeepLabv3[21] 和 ManTra-Net[22] 4 种方法在 CASIA v2 和 COLUMB 两个数据集上与本文方法进行比较,如表 3 和 4 所示。ManTra-Net 基于提取图像操作痕迹检测伪造区域,将所有的篡改类型集中在 1 个网络中检测,有一定的通用性。DeepLabv3 引入多孔空间金字塔池化结构,利用膨胀卷积扩展网络的感受野,从而高效提取不同尺度的上下文特征,能够提高模型对图像的识别精度。表 3 和 4 分别展示了所提方法在两个数据集上检测的数值,分析其数据可以看到,本文提出的方法在 Precision、Recall 和 $F1$ 均高于其他方法,说明了该方法具有一定的精度和可靠性。

表 3 CASIA v2 数据集对比实验

Table 3 Comparative experiments on the CASIA v2 dataset

方法	Precision	Recall	$F1$
ManTra-Net[22]	0.411 5	0.298 4	0.345 9
DeepLabv3[21]	0.832 5	0.722 5	0.773 6
U-Net[15]	0.787 9	0.655 1	0.715 4
RRU-Net[14]	0.868 4	0.721 9	0.788 4
本文	0.912 7	0.891 1	0.901 7

表 4　COLUMB 数据集对比实验

Table 4　Comparative experiments on the COLUMB dataset

方法	Precision	Recall	$F1$
ManTra-Net[22]	0.548 1	0.498 7	0.522 2
DeepLabv3[21]	0.925 4	0.884 6	0.904 5
U-Net[15]	0.867 3	0.795 4	0.829 8
RRU-Net[14]	0.946 1	0.894 2	0.919 4
本文	0.989 6	0.987 9	0.987 8

为了更直观地看出不同方法的检测能力，检验模型的篡改定位效果，从 CASIA v2 数据集中挑选 4 张具有代表性的小目标或含有微小区域的篡改图像和 1 张原始图像进行检测，如图 6 所示。其中，前 4 列图片为篡改图片，最后一列图片为原始图像(original image，OI)，第 2 行是掩码图像，白色区域是真实篡改区域，原始图像无掩码图像，下面的每一行表示每一种检测模型的检测结果。从图 6 中可以看出，由于 ManTra-Net 并未针对各类篡改类型专门建模，检测效果不佳；DeepLabv3 虽然具备基本的定位能力，但其定位精度有待提升；U-Net 虽能识别篡改区域，但在真实图像上易产生误判；RRU-Net 在性能上有所改进，但仍存在小尺度篡改区域漏检和原始图像误判的问题。相较而言，本文提出的检测方法展现出显著优势，不仅在原始图像上实现了零误判，而且对于篡改区域实现精准定位，特别是小目标区域和细节篡改区域也能够准确识别和定位，表现出模型优越的检测性能。

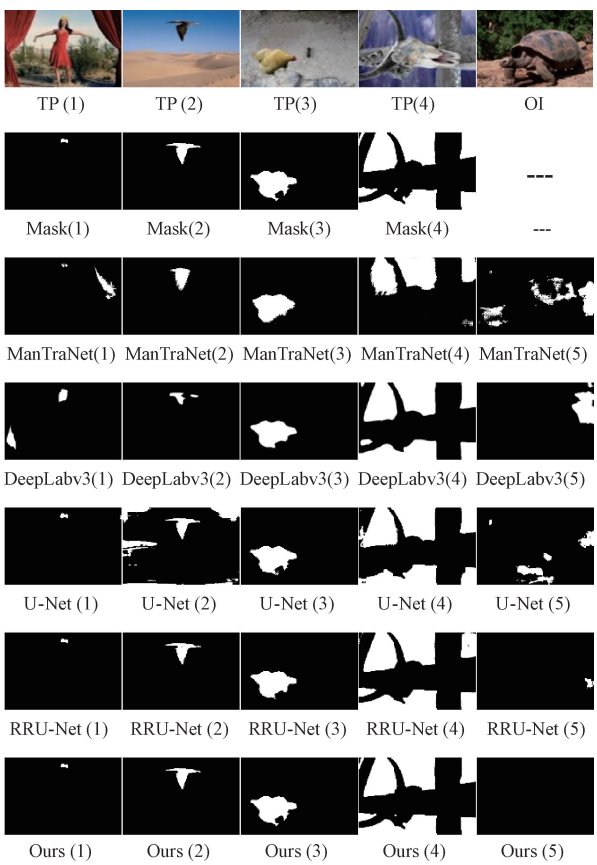

图 6　对比实验

Fig. 6　Comparative experiments

4　结　　论

本研究针对现有图像拼接篡改检测方法在细节与小目标定位方面的不足，提出了一种基于多尺度跨层融合的图像篡改定位模型。首先，设计了 MSC-SC 结构，引入 ASPP 处理编码器特征，通过 CLF 模块实现编码器细节特征与解码器语义特征的自适应融合，提升了模型对小目标和细节篡改区域的定位能力；其次，在编码器下采样路径中引入三重注意力机制，增强了模型对篡改区域的特征感知能力；最后，提出了 DBL 联合损失函数，提升了模型的训练稳定性和检测鲁棒性。由于在本文中数据集的分辨率不高，未来可以在更高分辨率的图像上实验研究，提高模型的鲁棒性和实用性。

参考文献

[1]　朱新同，唐云祁，耿鹏志. 数字图像篡改检测技术综述[J]. 中国人民公安大学学报(自然科学版)，2022，28(4)：87-99.
ZHU X T，TANG Y Q，GENG P ZH. Survey on digital image tampering detection technology [J]. Journal of People's Public Security University of China (Science and Technology)，2022，28(4)：87-99.

[2]　李昊东，庄培裕，李斌. 基于深度学习的数字图像篡改定位方法综述[J]. 信号处理，2021，37(12)：2278-2301.
LI H D，ZHUANG P Y，LI B. A survey on deep learning based digital image tampering localization methods [J]. Journal of Signal Processing，2021，37(12)：2278-2301.

[3]　PENG B，WANG W，DONG J，et al. Optimized 3D lighting environment estimation for image forgery detection [J]. IEEE Transactions on Information Forensics and Security，2016，12(2)：479-494.

[4]　杨超，周大可，杨欣. 基于篡改区域轮廓的图像拼接篡改盲取证算法[J]. 电子测量技术，2020，43(4)：132-138.
YANG CH，ZHOU D K，YANG X. Blind forensics for image tampering based on tampered contour[J]. Electronic Measurement Technology，2020，43(4)：132-138.

[5]　仲林林，吴奇，叶俊杰，等. 基于元学习的变电设备小样本缺陷图像检测[J]. 仪器仪表学报，2024，45(10)：154-167.
ZHONG L L，WU Q，YE J J，et al. Meta-learning-based few-shot image detection of defects in substation

equipment[J]. Chinese Journal of Scientific Instrument, 2024, 45(10): 154-167.

[6] TAN X G, ZHANG G M. Research on surface defect detection technology of wind turbine blade based on UAV image[J]. Instrumentation, 2022, 9(1): 41-48.

[7] RAO Y, NI J Q. A deep learning approach to detection of splicing and copy-move forgeries in images[C]. IEEE International Workshop on Information Forensics and Security, 2016: 1-6.

[8] 吴鹏, 陈北京, 郑雨鑫, 等. 基于双流 Faster R-CNN 的像素级图像拼接篡改定位算法[J]. 电子测量与仪器学报, 2021, 35(4): 154-160.
WU P, CHEN B J, ZHENG Y X, et al. Pixel-level image splicing localization algorithm based on dual-stream Faster R-CNN [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(4): 154-160.

[9] 周兴超, 魏为民, 裴仁莹, 等. 基于边缘特征增强的多尺度监督图像篡改检测[J]. 国外电子测量技术, 2023, 42(9): 122-129.
ZHOU X CH, WEI W M, PEI R Y, et al. Multi-scale supervised image tamper detection based on edge feature enhancement[J]. Foreign Electronic Measurement Technology, 2023, 42(9): 122-129.

[10] 张雷, 王宝丽, 陆晓栋, 等. 一种基于多尺度内容感知的图像篡改定位方法[J]. 山西大学学报(自然科学版), 2026, 49(2): 209-219.
ZHANG L, WANG B L, LU X D, et al. A multi-scale content-aware localization method for image manipulation [J]. Journal of Shanxi University (Natural Science Edition), 2026, 49(2): 209-219.

[11] 喻小芹, 单武扬, 邱骏颖, 等. 亮度对比度扰动下的图像篡改定位检测网络[J/OL]. 计算机应用, 1-13 [2026-04-27]. https://link. cnki. net/urlid/51. 1307. TP. 20250929. 1010. 004.
YU X Q, SHAN W Y, QIU J Y, et al. Image tampering localization and detection network under brightness-contrast disturbances [J/OL]. Journal of Computer Applications, 1-13[2026-04-27]. https://link. cnki. net/urlid/51. 1307. TP. 20250929. 1010. 004.

[12] CHEN L W, ZHU T G, WENG SH W, et al. A newly-constructed training dataset and an extremely-simplified network for copy-move forgery detection [J]. Expert Systems with Applications, 2026, 297: 129349.

[13] SHARMA K A, TIWARI R, DIXIT R, et al. Modelling of features of fusion using a hybrid swarm optimization algorithm with deep learning methodology for copy-move image forgery detection [J]. International Journal of System Assurance Engineering and Management, 2026, 17(1): 286-306.

[14] BI X L, WEI Y, XIAO B, et al. RRU-Net: The ringed residual U-Net for image splicing forgery detection [C]. IEEE/CVF Conference on Computer Vision and Pattern, 2019: 30-39.

[15] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.

[16] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.

[17] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module [C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3139-3148.

[18] HUANG G, LIU ZH, VAN DER MAATEN L. Densely connected convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[19] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510-519.

[20] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision (ECCV), 2018, 11211: 3-19.

[21] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. ArXiv preprint arXiv:1706. 05587, 2017.

[22] WU Y, ABDALMAGEED W, NATARAJAN P. ManTra-Net: Manipulation tracing network for detection and localization of image forgeries with anomalous features [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9543-9552.

作者简介

郭仕佳,硕士研究生,主要研究方向为图像篡改检测。
E-mail:hebut_114@163. com

张宝林(通信作者),博士,副教授,主要研究方向为信息安全、计算机视觉。
E-mail:tjutju@163. com

马琬雲,硕士研究生,主要研究方向为视频篡改检测。
E-mail:19245386259@163. com

张雨初,硕士研究生,主要研究方向为智能信息处理。
E-mail:jxufemwy@163. com