DOI:10.11918/202404005

面向图像修复的桥式注意力取证网络

张 澜,朱新山,王泽平,薛俊韬

(天津大学 电气自动化与信息工程学院,天津 300072)

摘 要:为提升多媒体信息的可靠性,减轻图像伪造事件对于社会造成的负面影响,亟需发展图像修复取证技术,检测并定位图像的篡改区域。本研究提出了一种面向图像修复的桥式注意力取证网络,该网络直接接收篡改后的图像,端到端的输出图像中被篡改的区域,网络采用编码器-解码器架构作为基础框架。首先,编码器选用 Swin Transformer 和 RepVGG 两个主干网络以提取多域修复特征。然后,使用桥式注意力模块连接两个主干网络的同级阶段,来增加编码器在局部和全局维度上的建模能力。最后,在编码器和解码器中间搭建了语义对齐融合模块,消除了两个主干网络提取的特征之间的语义不一致,有助于提高网络的取证性能。在不同修复取证数据集上的实验结果表明,所提出的模型与其他主流取证模型相比,能够更准确地对修复区域进行定位。特别是在有挑战性的 DeepFillV2 数据集和 Diffusion 数据集上,所提出的 BAFNet 分别取得了 91.37% 和 82.34% 的 IoU 分数,相比于主流的取证网络 MVSS-Net, IoU 指标分别提升了 8.77% 和 10.46%。另外,综合多个实验结果,BAFNet 在取证性能和模型复杂度之间取得了很好的平衡。

关键词:图像修复取证;深度取证网络;操作痕迹;多域修复特征;桥式注意力;语义对齐融合

中图分类号:TN911.73 **文献标志码**:A **文章编号**:0367-6234(2025)04-0062-09

Bridge-type attention forensics network for image inpainting

ZHANG Lan, ZHU Xinshan, WANG Zeping, XUE Juntao

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: To enhance the reliability of multimedia information and mitigate the negative impact of image forgery events on society, there is an urgent need to develop image inpainting forensics to detect and locate tampered regions of images. This paper proposes a bridge-type attention forensics network (BAFNet) for image inpainting. The network receives tampered images directly and outputs the tampered regions end-to-end. The network adopts an encoder-decoder architecture as the basic framework. Firstly, the encoder selects two backbones, Swin Transformer and RepVGG, to extract multi-domain inpainting features. Then, a bridge-type attention module is used to connect the same-level stages of the two backbones, enhancing the encoder's modeling capability in both local and global dimensions. Finally, a semantic alignment fusion module is built between the encoder and the decoder to eliminate semantic inconsistencies between the features extracted by the two backbones, thereby improving the forensic performance of the network. Experimental results on different inpainting forensic datasets demonstrate that the proposed model, compared with other mainstream forensic models, can more accurately locate the inpainting areas. In particular, on the challenging DeepFillV2 dataset and Diffusion dataset, the proposed BAFNet achieves IoU scores of 91.37% and 82.34%, respectively, which improves the IoU metrics by 8.77% and 10.46% compared to the mainstream forensic network MVSS-Net. In addition, combining the results of several experiments, BAFNet achieves a good balance between forensic performance and model complexity.

Keywords: image inpainting forensics; deep forensic network; manipulation traces; multi-domain inpainting features; bridge-type attention; semantic alignment fusion

随着图像编辑技术的快速发展,数字图像伪造的现象也越来越多[1-2]。例如:不法分子通过图像编辑技术伪造图像,误导各大媒体发布虚假新闻;一些人利用该技术制作虚假证据,误导法律机构以诽谤他人;此外,还有人使用图像编辑技术还原模糊图像,侵犯他人隐私等。近年来,人们越来越关注多媒体信息安全问题。对于开展检测并定位图像篡改区域的研究至关重要,越来越多的学者开始关注于图

收稿日期:2024-04-01;**录用日期**:2024-08-01;**网络首发日期**:2025-03-10
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20250310.1658.002
基金项目:国家自然科学基金(61972282,61971303)
作者简介:张 澜(2000—),男,硕士研究生;朱新山(1977—),男,教授,博士生导师
通信作者:薛俊韬,xuejt@tju.edu.cn

像篡改取证方向[3-5]。

图像修复是一种常见且有效的图像编辑技术,其目的是通过图像已有的信息,以视觉上合理的方式移除某部分内容或者填充修补缺失区域[6]。由于图像修复具有强大的图像编辑能力,与其他一些图像篡改操纵相比(比如复制 - 粘贴、拼接等),修复过程更加复杂,修复痕迹更加隐蔽,因此对于该技术的取证也更加具有难度。

图像修复取证技术可以分为两类:一类是基于传统方法的图像修复取证;另一类则是基于深度学习的图像修复取证。传统修复取证方法[7-11]主要的思路是将图像按区域划分为图像块,然后对图像块提取不同层次的手工特征,进一步检测图像块之间的相似度来判断图像哪块区域被篡改过。这些基于传统手工特征的图像修复检测算法在一些方面存在一定的局限性,比如需要手动选择区域、只针对特定的图像修复技术、鲁棒性较差等。

近年来为提高检测效率和性能,增强算法的鲁棒性,基于深度学习的图像修复取证算法也在不断发展。起初的一些算法[12-14]将卷积神经网络(convolutional neural networks, CNN)引入图像修复取证,通过神经网络自动提取篡改痕迹,并对修复区域进行定位,但性能有待提高。后来的一些算法[15-18]针对图像修复问题的特点,先将篡改图像进行滤波预处理,将图像从 RGB 域转换到噪声域,使其暴露篡改痕迹,再针对这些痕迹设计神经网络以进行篡改区域的检测。为了进一步捕获更全面的修复特征,一些学者搭建了双流神经网络[19],同时从 RGB 域和噪声域提取特征,但对双流提取的特征只做了简单的通道拼接融合,取证性能仍有待提高。

此外还有一些通用图像篡改检测算法[20-22],通过神经网络来提取图像中存在的异常篡改特征来判断图像中的异常区域,这些算法虽然对任意篡改方法均能进行检测,但是面对图像修复篡改时,性能并不令人满意。

通过总结现有的图像修复取证的研究现状,本文发现图像修复取证存在以下问题:1)大部分图像修复取证技术只从单域提取修复特征。从原始图像的 RGB 域提取修复特征,或者从滤波预处理图像的噪声域提取修复特征。这样的修复取证技术提取的修复特征单一,限制了取证性能。2)少部分图像修复取证技术从多域提取修复特征,即选择使用双流主干网络,同时在 RGB 域和噪声域提取修复特征,但是这些方法的双流主干网络是完全独立的,缺少双流之间的信息互补,限制了取证性能。3)图像修复取证网络在对多域特征融合时的方法过于简单,没有考虑到两个分支中的语义差异,缺少融合时的特征对齐,限制了取证性能。

针对当前图像修复取证中存在的不足之处,本文提出了一种面向图像修复的桥式注意力取证网络(bridge-type attention forensics network, BAFNet)。搭建了由双流主干网络组成的编码器,在并行的双流主干网络之间引入跨分支的桥式注意力模块,并在编码器和解码器的中间搭建了语义对齐融合模块。本文提出方案与其他主流取证模型相比,能够更准确地对修复区域进行定位,且具有良好的鲁棒性。

1　桥式注意力网络

1.1　网络的整体结构

图像修复取证任务可以看作一个语义分割问题。对于输入的待检测图像,通过判断其中的每个像素点是否被篡改过,进而检测出被篡改的区域,因此本文方法参考了语义分割的常用框架,也就是采用编码器 - 解码器架构作为基础框架。本文提出的网络结构如图 1 所示,除去编码器和解码器外,中间加入了语义对齐融合模块,用于对编码器提取的双域特征进行融合处理,再送入解码器生成预测结果图。具体来说,输入大小为 256 × 256 × 3 的图像,经过 BAFNet 后会得到 256 × 256 × 2 的输出,每个像素点两个通道的值分别代表该像素点未被篡改/被篡改的概率。

1.2　编码器设计

图像在被修复篡改的过程中,需要生成新的像素来填充图像中的缺失区域,而这些新的区域往往会留下一些视觉可见的线索,比如与真实区域的对比度差异以及纹理细节的差异,这些差异存在于图像的 RGB 域。另外,经已有研究发现,图像修复篡改中引入的新元素在噪声分布方面与真实区域不同[22],可以通过滤波器将图像转换到噪声域,提取噪声特征。图像修复取证不同于传统的语义对象检测,其更加关注篡改过程中留下的伪影而不是图像内容,因此需要编码器学习更丰富的特征。经过上述分析,本文设计了由双流主干网络组成的编码器,其中一流直接用原始 RGB 图像作为输入,提取与对比度、纹理相关的修复特征;另一流接收经过滤波器处理后的噪声图像,提取噪声相关的修复特征。

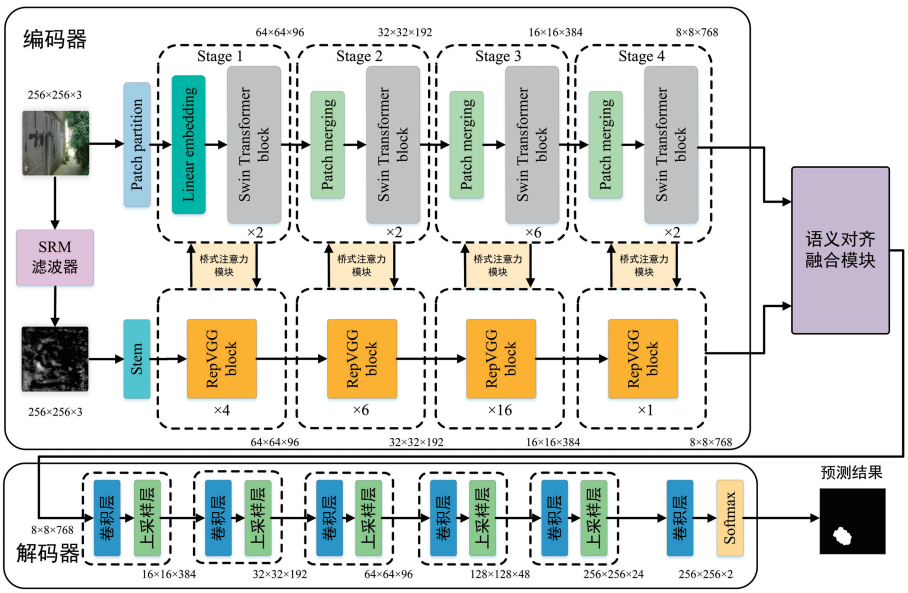

图 1 桥式注意力网络结构

Fig. 1 Architecture of bridge-type attention forensics network

如图 1 所示,编码器由双流主干网络组成,对于 RGB 流,篡改区域留下的纹理特征属于一种全局特征,需要提取图像的全局信息。考虑到 Transformer 善于长距离建模的优势,本文模型在 RGB 流选择使用 Transformer 从全局角度提取与修复特征相关的纹理细节。由于双流网络比传统的网络多一个主干网络,会引入加倍的计算开销,因此为了减少计算量,本文模型在 RGB 流选用 Swin Transformer[23] 的 Tiny 版本作为主干。该流共有 4 个阶段,对于 256 × 256 × 3 的输入图像,每个阶段的输出特征图大小分别为:64 × 64 × 96、32 × 32 × 192、16 × 16 × 384、8 × 8 × 768。

对于噪声流,需要先通过滤波器将原始 RGB 图像转换到噪声域。空域隐富模型(steganalysis rich model,SRM)[24]是一种用于数字图像隐写分析的高级模型,尤其擅长检测基于噪声的隐写方法。由于这个特点本文选用 SRM 来对图像进行滤波处理,经过 SRM 滤波后的图像可以暴露局部的噪声特征,这些特征更有利于网络去关注篡改留下的噪声信息,捕捉细节操纵伪影。如图 2 所示,本文模型采用文献[25]选取的 3 个 SRM 滤波核作为卷积核,输入 256 × 256 × 3 的 RGB 图像,映射得到同样尺寸的噪声域图像。考虑到卷积网络擅于检测局部细节的能力,可以从局部提取与修复特征相关的噪声信息,因此噪声流选用 RepVGG[26] 的 B0 版本作为主干,RepVGG 具有简单高效的架构,在保证了高性能的同时,可以通过结构重参数化技术减少计算开销,更适用于复杂的双流网络。同时本文模型调整了 RepVGG 每个阶段的输出通道数,保证与 Swin Transformer 各阶段输出特征图的形状相同。

$$\frac{1}{4}\begin{bmatrix}0&0&0&0&0\\0&-1&2&-1&0\\0&2&-4&5&0\\0&-1&2&-1&0\\0&0&0&0&0\end{bmatrix}\quad\frac{1}{12}\begin{bmatrix}-1&2&-2&2&-1\\2&-6&8&-6&2\\-2&8&-12&8&-2\\2&-6&8&-6&2\\-1&2&-2&2&-1\end{bmatrix}\quad\frac{1}{2}\begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\0&1&-2&1&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}$$

图 2 SRM 滤波核

Fig. 2 SRM filter kernels

1.3 桥式注意力模块

编码器中的双流主干网络在提取特征的过程中是完全独立的,本文模型在设计时认为双流主干各阶段提取的特征并非完全无关,噪声特征和 RGB 域纹理特征均属于修复篡改过程中留下的异常特征,如果通过设计模块将其连接,可以起到互相指导的作用,减少随着分辨率降低引起的与篡改相关信息的浪费。注意力机制允许模型在处理特征时,聚焦于最相关的信息,因此本文模型设计了桥式注意力模块连接双流主干的同级阶段,通过注意力机制提取与篡改区域最密切的特征,补充另一条主干网络。另外,分析编码器所选取的两个主干的特点,Transformer 更擅于远距离建模,因此 Transformer 的输出可以通过空间注意力的形式补充给卷积神经网络,进而帮助卷积神经网络感知特征图每个空间位置的重要性;卷积神经网络更擅于局部细节的检测,因此卷积神经网络的输出可以通过通道注意力的形式补充给 Transformer,帮助 Transformer 感知局部空间中每个通道的重要性。本文认为,双流提取的特征对于修复取证任务的作用是对等的,因此设计了双向的桥式注意力模块,如果仅选用单向的空间注

意力或者通道注意力,无法充分利用双流的特征,进而限制取证性能。

桥式注意力模块的结构如图 3 所示,Swin Transformer 每个阶段的输出特征图将经过空间注意力获取空间位置上的权重,然后与 RepVGG 同级阶段的输出特征图相乘。记 Swin Transformer 第 i 个阶段的输出特征图为 $\boldsymbol{T}_i$,计算空间位置上的权重矩阵的过程为

$$\boldsymbol{M}_s = \sigma(f^{7\times7}([\mathrm{AP}(\boldsymbol{T}_i);\mathrm{MP}(\boldsymbol{T}_i)])) \tag{1}$$

式中: $\boldsymbol{M}_s$ 为权重矩阵, $\sigma(\cdot)$ 为 Sigmoid 函数, $f^{7\times7}(\cdot)$ 为卷积核大小为 7 的卷积, $[\cdot]$ 为通道拼接, $\mathrm{AP}(\cdot)$、$\mathrm{MP}(\cdot)$ 分别为通道维度的平均池化(average pooling, AP)和最大池化(max pooling, MP)。然后与 RepVGG 第 i 个阶段的输出特征图 $\boldsymbol{C}_i$ 相乘,该过程可表示为

$$\boldsymbol{C}_i' = \boldsymbol{M}_s \otimes C_i \tag{2}$$

式中⊗为元素相乘,元素相乘过程中会将权重矩阵通过广播机制调整到与特征图维度相同。得到全局信息补充的特征图 C_i'将作为 RepVGG 第 $i+1$ 个阶段的输入。

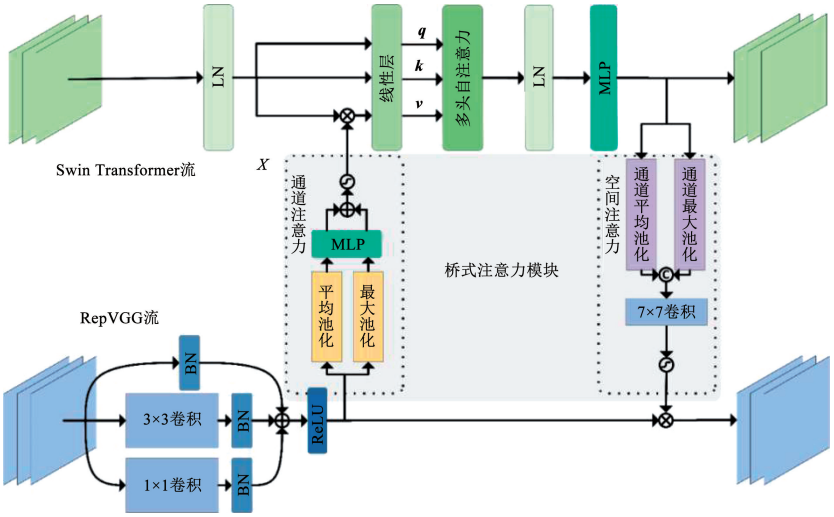

图 3　桥式注意力模块结构

Fig. 3　Architecture of bridge-type attention module

RepVGG 每个阶段的输出特征图将经过通道注意力获取局部空间各通道的重要性权重,然后与 Swin Transformer 同级阶段中第 1 个多头自注意力中生成 $\boldsymbol{v}$ 的特征向量相乘。首先计算通道维度的权重矩阵为

$$\boldsymbol{M}_c = \sigma(\mathrm{MLP}(\mathrm{AP}(\boldsymbol{T}_i)) + \mathrm{MLP}(\mathrm{MP}(\boldsymbol{T}_i))) \tag{3}$$

式中 $\mathrm{MLP}(\cdot)$ 为多层感知机。记多头自注意力机制的输入为 $\boldsymbol{x} \in \mathbf{R}^{N\times D}$,该输入是由输入特征图展平而来。$\boldsymbol{x}$ 在通过线性层生成 $\boldsymbol{v}$ 之前,将先与权重矩阵 $\boldsymbol{M}_c$ 相乘,改进后的多头自注意力机制中生成 $\boldsymbol{q}$、$\boldsymbol{k}$、$\boldsymbol{v}$ 的计算过程分别为:

$$\boldsymbol{q} = \boldsymbol{x}\boldsymbol{W}_q \tag{4}$$

$$\boldsymbol{k} = \boldsymbol{x}\boldsymbol{W}_k \tag{5}$$

$$\boldsymbol{v} = (\boldsymbol{M}_c \otimes \boldsymbol{x})\boldsymbol{W}_v \tag{6}$$

值得注意的是,为了减少计算量和模块复杂度,本文模型只改变了 Swin Transformer 每级阶段中的第 1 个多头自注意力。经过桥式注意力模块,可以将 Transformer 学习到的全局特征补充给卷积神经网络,增强卷积神经网络对于粗粒度特征的学习;并将卷积神经网络学习到的局部特征补充给Transformer,增强 Transformer 对于细粒度特征的学习。

1.4　语义对齐融合模块

已有双流主干网络对于两域信息的融合往往是采用简单的直接相加或者简单的通道维度拼接。这种融合方式忽略了两域中的语义差异,噪声域和 RGB 域特征的平等融合会导致语义的混乱,进而限制取证性能。基于这个考虑,本文设计了语义对齐融合模块。

语义对齐融合模块的结构如图 4 所示,其基本思路是对两域的特征进行加权平均。具体实现方案是对两域特征求和后先通过一种嵌入式通道注意力确定融合权重,然后使用该融合权重对两域特征进行加权融合。为了方便表示,记 CNN 流的输出特征为 $\boldsymbol{X}$,Transformer 流的输出特征为 $\boldsymbol{Y}$,经过语义对齐融合模块的输出特征为 $\boldsymbol{Z}$,则该模块的计算过程可表示为

$$\boldsymbol{Z} = \mathrm{ECA}(\boldsymbol{X}+\boldsymbol{Y})\otimes\boldsymbol{X} + (1-\mathrm{ECA}(\boldsymbol{X}+\boldsymbol{Y}))\otimes\boldsymbol{Y} \tag{7}$$

式中:$\mathrm{ECA}(\boldsymbol{X}+\boldsymbol{Y})$ 为计算得到的融合权重,$\mathrm{ECA}(\cdot)$ 为文献[27]所提出的嵌入式通道注意力。

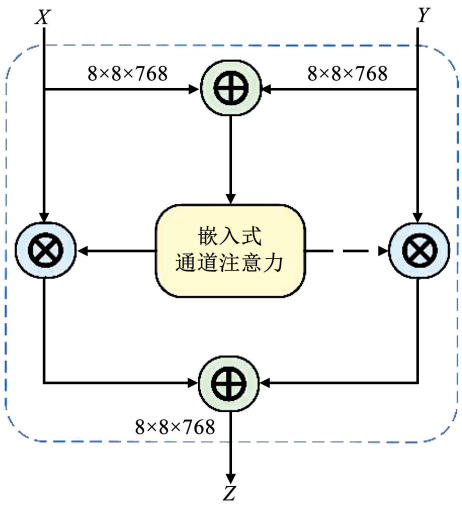

图 4 语义对齐融合模块结构

Fig. 4 Architecture of semantic alignment fusion module

嵌入式通道注意力通过在卷积操作中引入通道注意力机制,以捕捉不同通道之间的关系。结构如图 5 所示,其实现机制如下:首先通过全局平均池化层,将形状为 $H \times W \times C$ 的特征图压缩为 $1 \times 1 \times C$ 大小;然后计算自适应一维卷积的卷积核尺寸 k,计算公式为

$$k = \psi(C) = \left| \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right|_{\text{odd}} \tag{8}$$

式中:$b = 1$,$\gamma = 2$,C 为通道数,$|\cdot|_{\text{odd}}$ 是将一个不是奇数的数值向上舍入为最接近的奇数。最后使用自适应一维卷积处理压缩后的特征并经过 Sigmoid 函数,得到大小为 $1 \times 1 \times C$ 的融合权重向量。

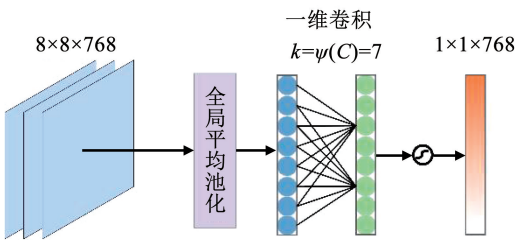

图 5 嵌入式通道注意力结构

Fig. 5 Architecture of embedded channel attention

选择嵌入式通道注意力作为融合权重的计算方式,可以让网络区分 RGB 域特征和噪声域特征各自关注的通道。语义对齐融合模块通过给特征分配不同权重来减少不同域特征在融合时引发的语义冲突,进而得到更有效的修复特征。

1.5 解码器设计

解码器的任务是将低分辨率的特征图恢复到原始分辨率,并生成像素级的分类结果。相较于编码器需要复杂的结构来提取输入图像的高级修复特征,解码器主要关注特征的空间重建和上采样,并不需要复杂的结构再对特征进行处理,因此解码器由卷积层和上采样层重复堆叠构成,结构见图 1。具体来说,卷积层由卷积核大小为 3,步长为 1 的卷积、BN 层、ReLU 激活函数组成,作用是将特征图的通道数减少 1/2,上采样层对特征图进行双线性插值,将特征图的尺寸增加 1 倍。卷积层和上采样层重复堆叠 5 次,将通道维度的信息逐渐转换到空间维度,每次堆叠后的输出特征图大小分别为 $16 \times 16 \times 384$、$32 \times 32 \times 192$、$64 \times 64 \times 96$、$128 \times 128 \times 48$、$256 \times 256 \times 24$。然后,经过一个卷积层调整特征图通道数为 2,再经过 Softmax 分类层得到最终的预测结果。

1.6 损失函数

为了更好地训练取证网络,需要设置一个合理的损失函数来优化网络参数。本文将修复取证任务看成一个逐像素分类的语义分割问题。另外考虑到在修复取证任务中修复篡改区域往往只占图片的一小部分,也就是说篡改像素点(正样本)远少于未篡改像素点(负样本),因此为了防止训练过程由像素数量多的类别所主导,缓解图像中存在的类别不均衡问题,本文模型采用了加权交叉熵损失,对正样本和负样本的损失赋予不同的权重,具体公式为

$$L = \frac{1}{N} \sum_{i=1}^{N} -[\omega_1 y_i \log p_i + \omega_2 (1 - y_i) \log(1 - p_i)] \tag{9}$$

式中:N 为像素点的总个数,y_i 为第 i 个样本,正样本取 1,负样本取 0;p_i 为第 i 个样本被预测为正样本的概率,ω_1、ω_2 分别为正样本和负样本的权重因子,本文在实验中取 $\omega_1 = 5$,$\omega_2 = 1$。

2 结果与分析

为了评估 BAFNet 在修复取证任务上的性能,本文使用 5 种图像修复算法构建了 5 个修复取证数据集。随后,BAFNet 在这些数据集上进行了训练和测试,并与一些典型的篡改取证方法进行了对比。此外,对于数据集采用了 JPEG 压缩、加噪的后处理操作,测试了网络的鲁棒性,最后进行了消融实验,验证了本文所提出的网络架构和组件的有效性。

2.1 数据集构建

本文在 Place 365[28] 数据集上随机挑选了 20 250张大小为 256×256 的 RGB 图像作为原始图像,Place 365 包含上百种场景类别,丰富的场景涵盖了图像篡改时可能涉及到的各种场景。图像修复算法可以大致分为:基于扩散的修复算法、基于样本的修复算法、基于深度学习的修复算法。为了更全面的评估本文提出的取证模型,在构建篡改数据集时要尽可能让数据集涵盖各类图像修复算法。具体

来说,在修复篡改过程中,本文挑选了 5 种典型图像修复算法,其中,有 3 种是近年来提出的基于深度学习的图像修复方法,分别是 DeepFillV2[29]、ICT[30]和 Lama[31]。其余两种分别是基于扩散的修复方法 Diffusion[32]和基于样本的修复方法 Exemplar[33]。构建数据集时首先在图像上随机去除一部分区域,随机包括位置的随机、形状的随机(包括圆形、矩形和不规则形状)、大小的随机。考虑到基于扩散的修复算法更适合较小的缺失区域,本文将基于扩散的修复算法的缺失区域大小设置为 0.10%、0.40%、1.56%、6.25%中的一种。其他方法的修复区域大小被设置为 1.00%, 5.00%, 10.00%中的一种。然后使用上述 5 种不同的修复方法对图像进行修复,得到 5 个不同的数据集(后文将用这 5 种修复方法的名字或缩写命名对应数据集)。最后,将每个数据集分为 3 个部分:18 000 张作为训练集,1 350 张作为验证集和 900 张作为测试集。

2.2　训练细节

本文基于 pytorch 框架构建了桥式注意力取证网络,并在 NVIDIA GeForce RTX 3090 GPU 上进行了训练。输入大小设置为 256×256×3,使用 Adam 优化器,批尺寸为 24,初始学习率设为 1×10^{-3},学习率采用余弦衰减。此外,为了防止网络过拟合,对图像进行了数据增强。数据增强包括随机水平或垂直翻转、随机方向旋转 90°和 JPEG 压缩,JPEG 压缩的质量因子在 70~100 之间随机选择。

为了与其他最先进的方法进行比较,本文选择了两种基于深度学习的修复取证方法,包括 FCNet[12]、IID-Net[16]。此外还选择了两种通用图像取证方法,包括 PSCC-Net[20]、MVSS-Net[22]。被比较的网络都严格按照对应论文中给出的训练程序和参数设置,加载预训练权重后并在本文的数据集上重新训练。

2.3　不同取证模型的定量比较

为了定量评估其他修复取证方法与本文所提出的方法,在 5 个数据集上对各个模型进行了训练和测试,并选择交并比(IoU)和 F1 分数作为评价指标。表 1 展示了每种取证方法的评估结果。

表 1　不同方案针对不同修复数据集的 IoU 和 F1 比较

Tab. 1　Comparison of IoU and F1 of different methods for different inpainting datasets　%

方法	Diffusion		Exemplar		ICT		Lama		DeepFillV2	
	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1
FCNet[12]	61.93	68.38	79.96	87.56	69.45	76.97	76.28	84.26	49.69	57.18
IID-Net[16]	59.97	64.29	82.63	89.63	76.66	84.30	81.40	88.75	54.42	61.92
PSCC-Net[20]	63.32	68.75	85.07	91.52	86.14	92.13	87.50	93.11	82.28	89.61
MVSS-Net[22]	71.88	77.64	88.79	93.82	86.52	92.33	87.04	92.79	82.60	89.61
BAFNet	**82.34**	**89.52**	**92.84**	**95.93**	**92.26**	**95.81**	**89.01**	**93.99**	**91.37**	**95.22**

由表 1 中结果可以直观看出,本文所提出的网络 BAFNet 在不同修复方法上制作的数据集上都取得了最佳的实验结果。比如在具有挑战性的取证数据集 DeepFillV2 和篡改区域面积较小的 Diffusion 数据集上,本文方法分别取得了 91.37%和 82.34%的 IoU 分数,比排在第 2 名的取证方法 MVSS-Net 提高了 8.77%和 10.46%。可以得出本文所提出的网络无论是面对基于深度学习修复方法制作的修复图像还是面对篡改区域较小的修复图像,相较于对比方案都有更显著的优势,尤其是相比于同样使用双流网络的 MVSS-Net,这也侧面证明了本文所提出的桥式注意力模块和语义对齐模块的优越性。

2.4　不同取证模型的定性比较

为了与其他方法进行更全面的比较,本文还对不同取证模型进行了定性评价比较(见图 6)。具体来说,展示了输入篡改图像(第 1 列)、篡改区域(第 2 列)以及不同取证方法(第 3 列~第 7 列)的检测结果。在前两行中展示了不同模型对于规则篡改区域的检测结果。虽然所有模型均可以对于篡改区域进行定位,但是与其他模型相比,本文模型可以准确地检测篡改区域的形状(圆形、正方形)。

在最后 3 行中展示了不同模型对于不规则篡改区域的检测结果。一般来说,不规则篡改区域的图像或者篡改区域面积小的图像对于取证任务是非常有挑战性的。但本文模型仍有不错的检测结果,比如图 6 中的第 3 行,不但可以定位到篡改区域,而且更好地拟合出了不规则的篡改边界。对于篡改面积特别小的图像,比如图 6 中后两行,部分对比方案出现了漏检的情况,而本文模型仍然检测出了篡改区域。并且与同样能检测出篡改区域的网络相比,具有更好的检测效果,形状方面更加拟合。可以看出对于不同情况下的篡改区域,BAFNet 均有不错的取证效果。

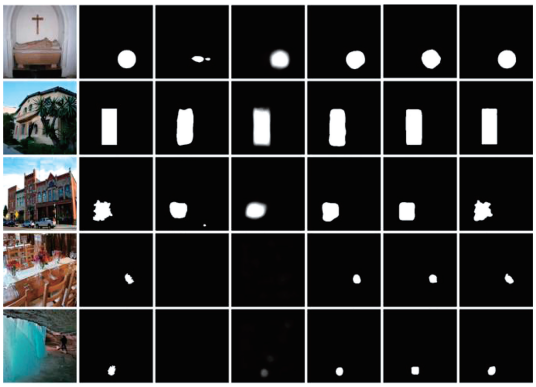

图 6 可视化取证结果

Fig. 6 Visualization of forensics results

2.5 不同取证模型的鲁棒性比较

在实际应用中,一些恶意的图像篡改者可能会通过后处理操作处理篡改图像,以隐藏篡改痕迹、逃避取证检测,因此本文对不同取证模型进行了鲁棒性比较。具体来说,在 DeepFillV2 数据集上对图像分别采用 JPEG 图像压缩和添加加性高斯白噪声(additive white Gaussian noise, AWGN)。选取的 JPEG 压缩的质量因子(Q_F)分别为 90、70, AWGN 的信噪比(R_{SN})分别为 50、40 dB。定量的评价结果见表 2。

表 2 不同方案在经过后处理的 DeepFillV2 数据集上的 IoU 和 F1 比较

Tab. 2 Comparison of IoU and F1 of different methods for the post-processed DeepFillV2 datasets %

后处理方法	FCNet[12]		IIDNet[16]		PSCC-Net[20]		MVSS-Net[22]		BAFNet	
	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1
JPEG 90	45.25	53.02	44.80	52.26	76.21	84.59	76.36	84.40	**88.25**	**92.94**
JPEG 70	42.17	49.69	29.39	36.32	51.02	59.28	50.45	59.00	**64.53**	**71.28**
加噪 50 dB	49.10	56.65	52.90	60.32	75.85	84.30	81.18	88.46	**91.02**	**95.08**
加噪 40 dB	46.82	54.53	45.56	52.83	39.06	49.99	81.38	87.90	**87.86**	**92.85**

可以看出,随着压缩系数和信噪比的降低,所有模型的取证性能都有所下降,而本文方法仍然是取证性能最好的方法。在具有挑战性的 DeepFillV2 数据集上,当 $Q_F=90$ 时,本文网络实现了 88.25% 的 IoU,当 $Q_F=70$ 时,本文网络实现了 64.53% 的 IoU。当 $R_{SN}=50$ dB 时,网络的 IoU 为 91.02%,当 $R_{SN}=40$ dB 时,网络的 IoU 为 87.86%。相比其他的网络,本文模型性能下降更少,对于后处理图像的检测性能甚至高于其他网络对于没有后处理的图像的检测性能,证明了本文模型具有很强的抗后处理能力。

2.6 网络面对未知修复方法的取证性能

在实际情况中,篡改图像的修复方法往往是未知的,这就要求修复取证模型能适用于未知的修复方法,因此本文评估了网络面对未知修复方法的取证性能,结果见表 3。具体来说,使用其中一种图像修复数据集训练模型,并使用其他图像修复数据集测试模型,即测试集中的修复方法对于模型训练是不可知的。值得注意的是,这个实验中没有选择 Diffusion 数据集,这是因为 Diffusion 数据集中修复区域的面积很小,这与其他数据集产生了领域差异。从实验结果可以看出,本文模型面对未知的修复方法时仍能表现出优越的取证性能。其中,在 DeepFillV2 数据集上训练的模型在跨数据集评估中表现出最优越的性能,而在其他数据集上训练的模型在 DeepFillV2 数据集上的评估效果较差。经过推断,这是由于 DeepFillV2 数据集对于修复取证任务更具挑战性。

评估结果表明,本文模型不是简单地提取与图像内容相关的语义特征,而是提取与修复信息相关的高级特征。

表 3 BAFNet 网络面对未知修复方法的 IoU 测试结果

Tab. 3 IoU for BAFNet against unknown inapinting methods

训练集	IoU/%				
	DeepFillV2	ICT	Lama	Exemplar	Mean
DeepFillV2	91.37	75.81	86.93	88.51	85.66
ICT	42.88	92.26	83.25	68.07	71.61
Lama	57.65	73.25	89.01	64.15	71.01
Exemplar	80.06	74.25	85.21	92.84	83.09

2.7 消融研究

本文设计了 5 个模型进行消融实验,来对所提出模型中各个模块的有效性进行验证:1)编码器为 RepVGG 的单流网络(模型Ⅰ);2)编码器为 Swin transformer 的单流网络(模型Ⅱ);3)编码器为 Swin transformer 和 RepVGG 的双流网络,但没有桥式注意力模块,特征融合部分为简单的通道拼接(模型Ⅲ);

4)编码器为 Swin transformer 和 RepVGG 的双流网络,双流中加入桥式注意力模块,语义融合部分为简单的通道拼接(模型Ⅳ);5)完整模型(模型Ⅴ)。5 个模型在 DeepFillV2 数据集上的表现见表 4,可以看出相较于模型Ⅰ和模型Ⅱ,使用双流网络从噪声域和 RGB 域同时提取修复特征的模型Ⅲ都有着超过 1.50% 的 IoU 性能提升。而加入桥式注意力模块的模型Ⅳ又有着 1.89% 的 IoU 性能提升,这证明了需要让双流网络在提取特征时进行信息互补。综合使用桥式注意力模块和语义对齐融合模块的完整模型Ⅴ又有 1.21% 的 IoU 性能提升,这说明将两域特征送入解码器前有必要进行特征对齐融合。综上所述,桥式注意力模块和语义对齐融合模块的引入,有效地提升了双流网络的取证性能。

表 4　在 DeepFillV2 数据集上的消融实验结果

Tab. 4　Ablation results on DeepFillV2 datasets

模型	RepVGG 单流	Swin transformer 单流	桥式注意力模块	跨域对齐融合模块	IoU/%	F1/%
模型Ⅰ	√				86.25	92.24
模型Ⅱ		√			86.65	92.50
模型Ⅲ	√	√			88.27	93.12
模型Ⅳ	√	√	√		90.16	94.54
模型Ⅴ	√	√	√	√	91.37	95.22

注:“√”表示在该行所代表的模型中包含对应结构或模块。

2.8　模型复杂度分析

本文使用了模型参数量和计算量(floating point operations,FLOPs)两个指标,对本文提出的模型和对比方案进行了复杂度评估,输入图像的尺寸被设置为 256 × 256 × 3,结果见表 5。从表 5 中可以看出,BAFNet 的计算量为 18.44 GFLOPs,参数量为 63.01 M。参数量高于单流网络,但远少于同为双流网络的 MVSS-Net,计算量最少。这表明本文网络在参数数量和计算量适中的情况下达到了最优性能,但仍有可优化的空间。

表 5　模型的复杂度对比

Tab. 5　Comparison of model complexity

模型	FLOPs/G	参数量/M
FCNet[12]	39.17	124.05
IID-Net[16]	33.25	5.78
PSCC-Net[20]	28.95	4.40
MVSS-Net[22]	40.90	142.79
BAFNet	18.44	63.01

3　结　论

1)设计双流主干网络作为整体网络的编码器,可以同时从噪声域和 RGB 域提取修复特征,丰富用于取证的特征。在双流主干网络中间搭建桥式注意力模块,有利于双流主干网络提取的特征之间的信息互补。

2)在编码器和解码器中间搭建语义对齐融合模块,对双流特征进行特征对齐融合,有效避免了特征融合时造成的语义混乱。

3)在 5 种图像修复方法制作的数据集上进行的实验,证明了提出的模型相较于对比方案,有着更好的取证效果以及鲁棒性,在取证性能和模型复杂度之间取得了很好的平衡。

参考文献

[1] ELHARROUSS O, ALMAADEED N, AL-MAADEED S, et al. Image inpainting: a review[J]. Neural Processing Letters, 2020, 51(2): 2007. DOI: 10.1007/s11063-019-10163-0

[2] VERDOLIVA L. Media forensics and DeepFakes: an overview[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(5): 910. DOI: 10.1109/JSTSP.2020.3002101

[3] 谢伟, 万晓霞, 叶松涛, 等. 基于局部色彩不变量的图像篡改检测方法[J]. 湖南大学学报(自然科学版), 2016, 43(8): 128
XIE Wei, WAN Xiaoxia, YE Songtao, et al. Image copy-move forgery detection based on local color invariants[J]. Journal of Hunan University (Natural Sciences), 2016, 43(8): 128. DOI: 10.16339/j.cnki.hdxbzkb.2016.08.019

[4] 孙鹏, 郎宇博, 樊舒, 等. 图像拼接篡改的自动色温距离分类检验方法[J]. 自动化学报, 2018, 44(7): 1321
SUN Peng, LANG Yubo, FAN Shu, et al. Detection of image splicing manipulation by automated classification of color temperature distance[J]. Acta Automatica Sinica, 2018, 44(7): 1321. DOI: 10.16383/j.aas.2017.c170267

[5] 朱新山, 卢俊彦, 甘永东, 等. 融合多尺度特征与多分支预测的多操作检测网络[J]. 湖南大学学报(自然科学版), 2023, 50(8):94
ZHU Xinshan, LU Junyan, GAN Yongdong, et al. Multi-manipulation detection network combining multi-scale feature and multi-branch prediction[J]. Journal of Hunan University (Natural Sciences), 2023, 50(8): 94. DOI: 10.16339/j.cnki.hdxbzkb.2023285

[6]BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting [C]//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques-SIGGRAPH'00. New York: ACM, 2000: 417. DOI: 10.1145/344779.344972

[7]WU Qiong, SUN Shaojie, ZHU Wei, et al. Detection of digital doctoring in exemplar-based inpainted images [C]//2008 International Conference on Machine Learning and Cybernetics. Kunming: IEEE, 2008: 1222. DOI: 10.1109/ICMLC.2008.4620591

[8]BACCHUWAR K S, Aakashdeep, RAMAKRISHNAN K R. A jump patch-block match algorithm for multiple forgery detection[C]//2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s). Kottayam: IEEE, 2013: 723. DOI: 10.1109/iMac4s.2013.6526502

[9]CHANG I C, YU J C, CHANG C C. A forgery detection algorithm for exemplar-based inpainting images using multi-region relation[J]. Image and Vision Computing, 2013, 31(1): 57. DOI: 10.1016/j.imavis.2012.09.002

[10]LIANG Zaoshan, YANG Gaobo, DING Xiangling, et al. An efficient forgery detection algorithm for object removal by exemplar-based image inpainting[J]. Journal of Visual Communication and Image Representation, 2015, 30: 75. DOI: 10.1016/j.jvcir.2015.03.004

[11]LI Haodong, LUO Weiqi, HUANG Jiwu. Localization of diffusion-based inpainting in digital images[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(12): 3050. DOI: 10.1109/TIFS.2017.2730822

[12]ZHU Xinshan, QIAN Yongjun, ZHAO Xianfeng, et al. A deep learning approach to patch-based image inpainting forensics[J]. Signal Processing: Image Communication, 2018, 67: 90. DOI: 10.1016/j.image.2018.05.015

[13]WANG Xinyi, WANG He, NIU Shaozhang. An image forensic method for AI inpainting using faster R-CNN [C]//Artificial Intelligence and Security. Cham: Springer International Publishing, 2019: 476. DOI: 10.1007/978-3-030-24271-8_43

[14]LU Ming, NIU Shaozhang. A detection approach using LSTM-CNN for object removal caused by exemplar-based image inpainting[J]. Electronics, 2020, 9(5): 858. DOI: 10.3390/electronics9050858

[15]LI Haodong, HUANG Jiwu. Localization of deep inpainting using high-pass fully convolutional network [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 8300. DOI: 10.1109/ICCV.2019.00839

[16]WU Haiwei, ZHOU Jiantao. IID-net: image inpainting detection network via neural architecture search and attention[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1172. DOI: 10.1109/TCSVT.2021.3075039

[17]任洪昊, 朱新山, 卢俊彦. 深度图像修复的动态特征融合取证网络[J]. 哈尔滨工业大学学报, 2022, 54(11): 47
REN Honghao, ZHU Xinshan, LU Junyan. Dynamic feature fusion forensics network for deep image inpainting[J]. Journal of Harbin Institute of Technology, 2022, 54(11): 47. DOI: 10.11918/202201081

[18]窦立云, 冯国瑞, 钱振兴, 等. 抗深度取证的多粒度融合图像修复网络[J]. 上海大学学报(自然科学版), 2023, 29(1): 10
DOU Liyun, FENG Guorui, QIAN Zhenxing, el al. Multi-granularity fusion-based image inpainting network resistant to deep forensics [J]. Journal of Shanghai University (Natural Science Edition), 2023, 29(1): 10. DOI: 10.12066/j.issn.1007-2861.2456

[19]ZHU Xinshan, LU Junyan, REN Honghao, et al. A transformer—CNN for deep image inpainting forensics[J]. The Visual Computer, 2023, 39(10): 4721. DOI: 10.1007/s00371-022-02620-0

[20]LIU Xiaohong, LIU Yaojie, CHEN Jun, 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. DOI: 10.1109/TCSVT.2022.3189545

[21]WU Yue, ABDALMAGEED W, NATARAJAN P. Mantra-net: manipulation tracing network for detection and localization of image forgeries with anomalous features [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 9535. DOI: 10.1109/CVPR.2019.00977

[22]DONG Chengbo, CHEN Xinru, HU Ruohan, 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. DOI: 10.1109/TPAMI.2022.3180556

[23]LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 9992. DOI: 10.1109/ICCV48922.2021.00986

[24]FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868. DOI: 10.1109/TIFS.2012.2190402

[25]ZHOU Peng, HAN Xintong, MORARIU V I, et al. Learning rich features for image manipulation detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1053. DOI: 10.1109/CVPR.2018.00116

[26]DING Xiaohan, ZHANG Xiangyu, MA Ningning, et al. RepVGG: making VGG-style ConvNets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 13728. DOI: 10.1109/CVPR46437.2021.01352

[27]WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 11531. DOI: 10.1109/CVPR42600.2020.01155

[28]ZHOU Bolei, LAPEDRIZA A, KHOSLA A, et al. Places: a 10 million image database for scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(6): 1452. DOI: 10.1109/TPAMI.2017.2723009

[29]YU Jiahui, LIN Zhe, YANG Jimei, et al. Free-form image inpainting with gated convolution [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 4470. DOI: 10.1109/ICCV.2019.00457

[30]WAN Ziyu, ZHANG Jingbo, CHEN Dongdong, et al. High-fidelity pluralistic image completion with transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 4672. DOI: 10.1109/ICCV48922.2021.00465

[31]SUVOROV R, LOGACHEVA E, MASHIKHIN A, et al. Resolution-robust large mask inpainting with Fourier convolutions [C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2022: 3172. DOI: 10.1109/WACV51458.2022.00323

[32]TSCHUMPERLÉ D, FOUREY S, OSGOOD G. G'MIC: an open-source self-extending framework for image processing. Journal of Open Source Software, 2025, 10(105): 6618. DOI: 10.21105/joss.06618

[33]CRIMINISI A, PÉREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting [J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200. DOI: 10.1109/tip.2004.833105

(编辑 张 红)