

融入混合注意力的低缩放因子 Seam Carving 篡改检测算法

赵洁, 常皓婵, 武斌

(天津城建大学计算机与信息工程学院, 天津 300384)

摘要: 针对现有的 Seam Carving 篡改检测算法对于低缩放因子情况存在检测精度不高、鲁棒性不强的问题, 提出一种融入混合注意力机制的 Seam Carving 篡改检测算法。首先, 利用 BayarConv2D 约束卷积对图像进行预处理, 充分学习图像的噪声特征, 并通过矩阵乘法与 RGB 图像进行特征融合; 然后, 采用 ResNet 作为骨干网络进行特征学习, 引入残差传播和残差反馈机制, 凸显 Seam Carving 的操作痕迹; 最后, 利用混合注意力机制同时提取相邻位置和通道之间的特征, 更好地捕捉全局特征, 进而将其输入全连接层进行分类。实验结果表明, 在 BOSSbase1.01 数据集上, 当缩放因子为 1% 和 9% 时, 检测精度分别达到了 89.48% 和 97.94%, 优于现有主流方法, 同时具有较低的计算复杂度和较好的鲁棒性, 能够抵抗 JPEG 压缩攻击。

关键词: 混合注意力机制; 图像取证; Seam Carving 检测; 低缩放因子

中图分类号: TP309.2

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202414

Low scaling factor Seam Carving tamper detection algorithm with hybrid attention

ZHAO Jie, CHANG Haochan, WU Bin

School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300384, China

Abstract: The existing seam carving tamper detection algorithms have the problems of low detection accuracy and weak robustness for the case of low scaling factor. A seam carving tamper detection algorithm integrated with hybrid attention mechanism is proposed. Firstly, BayarConv2D constrained convolution is used to preprocess the image, fully learn the noise characteristics of the image, and fuse the features with RGB image through matrix multiplication. Then, ResNet is used as the backbone network for feature learning, and the residual propagation and residual feedback mechanisms are introduced to highlight the operation traces of seam carving. Finally, the hybrid attention mechanism is used to simultaneously extract the features between adjacent locations and channels to better capture the global features, and then input them into the full connection layer to achieve classification. The experimental results show that on the BOSSBase1.01 dataset, when the scaling factor is 1% and 9%, the detection accuracy of the proposed method reaches 89.48% and 97.94% respectively, which is better than existing mainstream methods. At the same time, it has lower computational complexity and better robustness, and can resist JPEG compression attacks.

Key words: mixed attention mechanism, image forensics, Seam Carving detection, low scaling factor

收稿日期: 2024-03-15; 修回日期: 2024-05-22

通信作者: 武斌, wubin@tcu.edu.cn

基金项目: 天津市重点研发计划科技支撑重点项目 (No.19YFZCGX00130); 天津市企业科技特派员项目 (No.19JCTPJC47200)

Foundation Items: The Key Project Supported by Key Research and Development of Tianjin (No.19YFZCGX00130), The Enterprise Science and Technology Commissioner Project of Tianjin (No.19JCTPJC47200)

0 引言

数字图像取证技术, 通过分析数字图像的特征来对图像的真实性、完整性以及来源进行鉴别, 逐渐成为信息安全领域的研究热点之一。目前主要分为两类: 一类是检测对图像内容进行实质性改变的操作, 如图像拼接^[1]、复制粘贴^[2]和人脸伪造图像^[3]等; 另一类是检测不改变内容但改变图像特征和结构的操作, 如重采样^[4]、JPEG 压缩^[5]、中值滤波^[6]和图像重定向^[7]等。Seam Carving 图像缩放技术是一种常见的图像重定向方法。为了实现图像在不同种类和尺寸电子设备上的完美显示, 需要在不扭曲图像内容的同时实现图像尺寸的缩放。传统的图像缩放技术^[8]是基于重采样原理实现的, 该技术直接对图像进行拉伸、删改, 会严重破坏图像内容。为了尽量保留图片的重要区域、减少视觉失真, Shamir 等^[9]提出的一种基于内容感知的图像缩放技术——Seam Carving, 主要通过移除或插入图像中不重要的像素点来达到改变图像尺寸的目的。然而, 不法分子也会利用该技术对图像进行篡改, 通过人为缩小或移除图像中的某些重要目标来改变原始图像的语义内容。因此, 针对 Seam Carving 技术的篡改检测已经成为图像取证领域的一个研究热点。

目前存在的 Seam Carving 篡改检测方法可以分为传统检测方法和基于深度学习的检测方法两大类。传统检测方法一般首先手工提取图像处理前后的特征变化, 然后将其输入 SVM 中进行分类。Sarkar 等^[10]通过分析 Seam Carving 操作与隐写分析的相似性, 首次利用 Shi-324 统计特征检测 Seam Carving 操作。Wattanachote 等^[11]利用块效应特征矩阵 (blocking artifact characteristics matrix, BACM) 可以检测 JPEG 图片里面的 Seam Carving 操作。Yin 等^[12]观察到 Seam Carving 会引起图像的纹理失真, 提出了针对局部二值模式 (local binary patterns, LBP) 的检测算法。Senturk 等^[13]提出了一种逐块检测缝雕刻的方法, 将每个块转换成能量图, 然后直接提取基于缝的特征。现有的传统检测方法都没有针对低缩放因子的情况进行实验, 并且存在检测准确率不高、鲁棒性不强等问题。

近年来, 由于卷积神经网络 (convolutional neural network, CNN) 在计算机视觉的广泛应用, 许多学者将其引入图像篡改检测领域。Cieslak 等^[14]将图像转换到 LBP 域, 然后利用 VGG16 进行检

测。Ye 等^[15]提出了一种基于 CNN 的未压缩数字图像接缝雕刻检测模型, 准确率有所提升, 当缩放因子为 5% 时, 准确率已提升至 90.4%。受 SRNet^[16]的启发, Nam 等^[17]提出了一种基于 CNN 的取证方案, 其由 5 种类型的网络模块组成, 但是只针对缩放因子 10%~50% 的图像进行实验, 未提及低缩放因子情况。Nataraj 等^[18]提出了一种两阶段模型, 阶段 1 定位, 阶段 2 执行分类, 通过生成热图来定位图像中的接缝雕刻, 但该方法准确度较低, 当缩放因子为 1% 时, 准确率仅有 64.6%。Moreira 等^[19]提出了一种改进 Xception 架构的方法来解决接缝雕刻检测的问题, 该算法缩放因子为 3% 时, 准确率为 85.2%。Xia 等^[20]引入减法像素邻接矩阵 (subtractive pixel adjacency matrix, SPAM) 和空间金字塔池化的检测网络, 当缩放因子为 1%~9% 时, 准确率为 86.1%~96.2%。

目前, 无论是传统方法还是基于深度学习的方法, 在低缩放因子情况 (即 1%~9%) 下, 都存在检测准确率不高的问题, 特别在缩放因子低至 1% 时准确率明显下降, 且模型鲁棒性不高, 难以抵抗 JPEG 压缩攻击。针对上述问题, 本文提出了一种引入混合注意力机制的 CNN 模型, 有效提高对低缩放因子情况的检测精度, 并且对 JPEG 压缩具有较强的鲁棒性。本文工作的主要贡献总结如下。

(1) 为了充分考虑篡改图像中的噪声特征, 引入 BayarConv2D^[21]进行噪声提取, 该噪声提取方式优于 SRM 滤波器, 能够抑制图像的语义信息。将经过 BayarConv2D 处理后的图像与 RGB 图像进行特征融合, 作为 CNN 的输入。

(2) 为了强化 CNN 的学习方式, 受 RRU-Net^[22]的启发, 在网络中引入残差传播和残差反馈机制。残差传播对输入特征信息进行回忆, 能解决深层网络中的梯度退化问题; 残差反馈对输入特征信息进行整合, 使 Seam Carving 操作痕迹更加明显。

(3) 引入混合注意力模块, 由位置注意模块和通道注意模块两个模块并联组成, 能同时提取相邻位置和通道之间的特征, 将两个模块的输出聚合起来, 以获得更好的特征表示, 用于像素级预测。

1 Seam Carving 算法原理

Seam Carving 是一种基于内容感知的图像重定向方法, 通过不断去除最低能量像素接缝来调整图像大小。像素点的能量函数定义如式 (1) 所示。

$$e(i, j) = \left| \frac{\partial I}{\partial x} \right| + \left| \frac{\partial I}{\partial y} \right| \quad (1)$$

以竖直方向为例, seam是指从图像顶端到底端由八连通的像素点连接起来的位置路径。对于尺寸为 $N_1 \times N_2$ 的图像 I , seam的定义可以用式(2)来表达。其中 S 代表一条 seam所包含的像素集合, x 代表从 $[1, \dots, N_1]$ 到 $[1, \dots, N_2]$ 的映射, 参数 i 表示 seam 中第 i 个像素点的横坐标, $x(i)$ 表示 seam 中第 i 个像素点的纵坐标。

$$S^x = \{S_i^x\}_{i=1}^{N_1} = \{i, (x(i))\}_{i=1}^{N_1}, |x(i) - x(i-1)| \leq 1 \quad (2)$$

为了寻找最优 seam, 首先利用动态规划的思想建立最小累积能量矩阵 M , 如式(3)所示, 其中 $i \in [1, N_1]$, $j \in [1, N_2]$ 。然后根据最后一行最小值, 利用八连通路径进行反向遍历, 得到最优的 seam, 即待删除像素点的路径位置。

$$M(i, j) = \min \{M(i-1, j-1), M(i-1, j), M(i-1, j+1)\} + e(i, j) \quad (3)$$

以图1为例, 在原始图像的基础上利用 Seam Carving 技术进行缩放, 图1(a)为原始图像, 图像主要部分是球场上的人。图1(b)和图1(c)分别为缩放因子为5%和30%时的图像, 图1(d)和图1(e)分别为被恶意篡改的图像。通过观察可以发现, 缩放因子越小, 篡改痕迹越不明显, 检测难度越大。

2 网络结构

2.1 网络总体结构

本文提出了一种融入混合注意力机制的 Seam Carving 篡改检测算法, 网络结构如图2所示, 由4个部分组成, 包括预处理模块、特征学习模块、混合注意力模块以及输出模块。预处理模块采用 BayarConv2D 约束卷积来抑制图像内容, 以突出缝刻伪影。特征学习模块采用 ResNet^[23] 作为骨干网, 引入残差传播和残差反馈机制, 以解决深层网络中

的梯度退化问题, 充分学习图像特征。混合注意力机制能同时提取相邻位置和通道之间的特征, 提高识别准确率。最后的输出模块由典型的全连接层和 softmax 函数组成, 是一个线性分类器。在该网络中, 除第一层采用 5×5 核外, 其余卷积层均采用 3×3 核, 所有非线性激活函数均采用整流线性单元 (rectified linear unit, ReLU)。该网络中支持两种具有快捷方式的块, 如图2中的 Block2 和 Block4, Block2 由于特征维数相同, 直接连接输入和输出, Block4 使用 1×1 卷积, 步长为2, 执行快捷方式匹配两个不同维数的特征图。图2中 A 、 B 表示输入特征图, C 表示特征图的通道数, W 和 H 分别表示特征图的宽和高。

2.2 图像预处理模块

图像取证不同于经典的图像分类和识别任务, 它在很大程度上依赖于独立于图像内容的细微伪影或痕迹^[21]。因此, 大多数现有的取证工作^[24-26]并没有直接从原始图像中学习特征, 通常进行预处理, 如利用频率分析来抑制图像内容, 突出图像操作伪影, 以更好地学习特征。本文提出利用 BayarConv2D 作为检测算法的预处理方法, BayarConv2D 是一个约束卷积层, 滤波器的中心值固定为-1, 其他值的和固定为1, 所有权重均归一化为0。设滤波器的权重为 ω_{ij} , 权重约束如式(4)所示:

$$\begin{cases} \sum_i \sum_j \omega_{ij} = 0 \\ \omega_{0,0} = -1 \end{cases} \quad (4)$$

2.3 残差传播与残差反馈

本文提出的网络采用 ResNet 作为骨干网, 其中的快捷连接模块可以帮助梯度向上传播, 有效缓解训练过程中梯度消失的问题。受到 RRU-Net^[22] 的启发, 本文在 CNN 中引入残差传播和残差反馈, 具体结构如图3所示。

残差传播对输入特征信息进行回忆, 可以解决深层网络中的梯度退化问题, 残差映射定义为式(5):

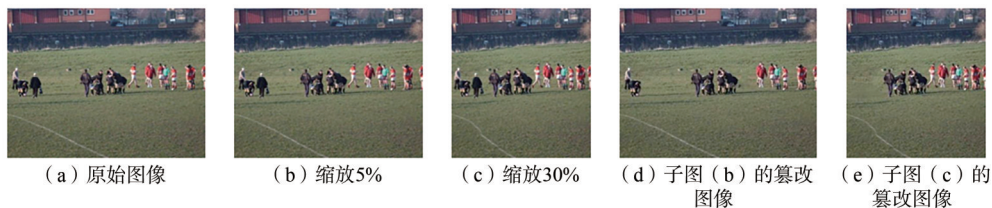


图1 Seam Carving 图像缩放实例

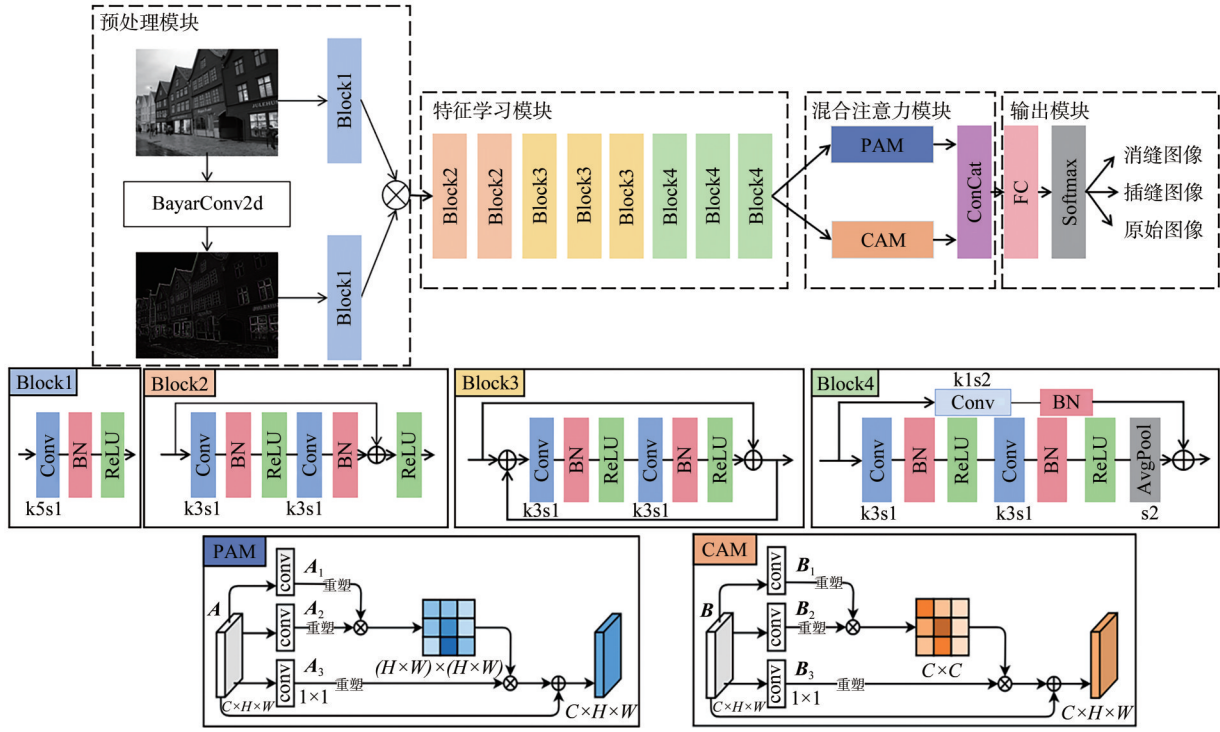


图2 本文算法的网络结构

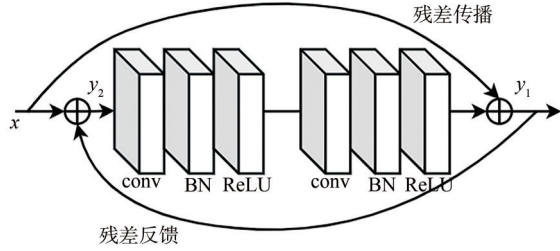


图3 残差传播和残差反馈结构图

$$y_1 = F(x) + x \quad (5)$$

其中, x 为输入, y_1 为输出, $F(x)$ 表示卷积输出结果, $F(x) + x$ 表示运用了快捷连接和元素加法。

残差反馈对输入特征信息进行整合, 采用带有 sigmoid 激活函数的门控机制来学习判别特征通道之间的非线性相互作用, 再将 sigmoid 得到的响应值叠加在输入 x 上, 并与输入 x 相加, 使 Seam Carving 操作痕迹更加明显, 反馈函数定义为式 (6):

$$y_2 = (s(I(y_1)) \times x) + x \quad (6)$$

其中, x 为输入, y_1 为残差传播的输出, y_2 为残差反馈的输出, $I()$ 为一个线性映射, 用于改变 y_1 维度, $s()$ 为 sigmoid 激活函数。

2.4 混合注意力机制

为了更好地捕捉全局特征, 本文在网络中设计了混合注意力机制, 具体结构如图4所示。

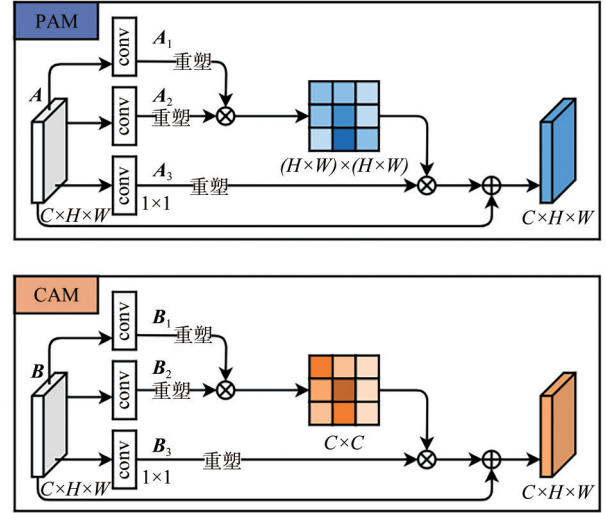


图4 混合注意力机制结构

如位置注意力模块 (position attention module, PAM) 所示, 首先输入特征 A 利用卷积层获得了3个降维特征 A_1 、 A_2 和 A_3 。然后通过以下3个步骤生成空间远程上下文信息的新特征, 首先是对特征 A_1 和 A_2 之间的空间关系进行建模, 生成空间注意矩阵 F , 其中:

$$F_{ji} = \frac{\exp(A_{1i} \cdot A_{2j})}{\sum_{i=1}^N \exp(A_{1i} \cdot A_{2j})} \quad (7)$$

其中, F_{ji} 为位置 i 对位置 j 的影响, 两个位置的特

征表示越相似，它们之间的相关性越大。

接下来，在注意矩阵 F 和特征 A_3 之间执行矩阵乘法。最后，对上述相乘的结果矩阵和原始特征执行元素求和操作，以获得长期上下文的最终表示，最终输出如式 (8) 所示：

$$E_j = \alpha \sum_{i=1}^N (F_{ji} \cdot A_{3i}) + A_j \quad (8)$$

其中， α 为初始化是 0、可学习的参数。

通道注意模块 (channel attention module, CAM) 捕获通道维度上的远程上下文信息。捕获通道关系的过程类似于位置注意模块，除了第一步需要在信道维度中计算信道注意矩阵。最后，将两个关注模块的输出聚合起来，以获得更好的特征表示，用于像素级预测^[27]。

3 实验结果及分析

为了充分评估本文提出的网络模型对低缩放因子 Seam Carving 操作痕迹的检测性能，我们进行了一系列实验和分析。本节提供了实验的设置细节和结果分析。

3.1 数据集

实验采用 BOSSbase1.01 数据集^[28]生成基准数据集，包含 1 万张大小为 512×384 的灰度原始图像。与 ILFNet^[17]类似，使用质量系数为 100 的 JPEG 压缩保存获得的原始图像。基于 Seam Carving 算法，对原始图像进行以下操作。

- 进行低缩放因子垂直消缝操作和垂直缝插入操作，从 1% 到 9%，步长为 1%，即比例为 1%、2%、3%、4%、5%、6%、7%、8% 和 9%，共获得 9 万张消缝图片和 9 万张插缝图片。

- 进行高缩放因子垂直消缝操作和垂直缝插入操作，从 10% 到 50%，步长为 10%，即比例为 10%、20%、30%、40%、50%，共获得 5 万张消缝图片和 5 万张插缝图片。

- 与原始图像一样，生成的消缝和插缝图像也使用质量因子为 100 的 JPEG 压缩保存。

此时，我们总共获得了 29 万张分辨率在 256×384 和 768×384 之间的图像。将上述得到的图像数据集按照 8:1:1 的比例分成 3 组进行训练、验证和测试。在训练和测试过程中，每个类别对应的图像数据的比例保持在 1:1:1。在输入网络之前，采用随机裁剪策略将图像裁剪为 256×256。

3.2 训练参数设置

采用 PyTorch 框架构建网络，并使用 NVIDIA GeForce RTX 3080 Ti 训练模型。经过反复实验，训练过程使用学习率为 10^{-3} 的 Adam 优化器，动量系数 $\beta_1=0.9$ 和 $\beta_2=0.99$ ，数值稳定常数 $\epsilon=10^{-8}$ 。Batch-size 大小设置为 16。学习率每经过 10 轮迭代进行一次递减，总共训练 100 轮。

3.3 评估指标

为了比较算法之间的优劣，实验必须采取客观的评价指标。本文主要从两个方面对检测算法进行评价：一个是检测准确率，另一个是检测算法的鲁棒性。

- 检测准确率：在给定的图像测试集中，分类器根据图像特征能够正确进行分类的样本数与总样本数之比就是检测准确率，常用百分数表示。在本文实验中，主要是衡量算法能够正确区分原始图像、消缝图像和插缝图像的能力，其判定过程属于一个三分类问题。

- 检测算法的鲁棒性：在实际检测过程中，图像的篡改程度未知，具有鲁棒性的算法应该能够在篡改程度不同的图像测试集中取得稳定的检测效果。同时，具有鲁棒性的检测算法应该能够有效抵抗 JPEG 压缩攻击。

3.4 消融实验

为了验证本文网络模型中各个模块的有效性，我们对模块进行了消融实验。如前文所述，本文提出的网络由 7 个模块组成，包括 BayarConv2D、Block1、Block2、Block3、Block4、PAM 和 CAM。消融实验结果如表 1 所示，检测准确率表示缩放因子为 1%~50% 的混合测试集的性能。与单独使用 BayarConv2D、Block2、Block3 和 Block4 模块相比，组合使用这些模块可以实现最佳的模型性能。使用 BayarConv2D 模块，能够充分提取图像的噪声特征，在 Block3 模块中引入其残差传播和残差反馈，在 Block2 和 Block4 中引入残差连接，与 Block3 组合使用能够提高模型特征学习能力并且解决梯度退化问题。在此基础上加入注意力机制，组合使用 PAM 和 CAM，即分别在空间维度和通道维度上对语义相互依赖进行建模，能够提高模型检测准确率。

图 5 展现了部分消融实验的可视化效果，样本为 3 000 张缩放因子为 1%~50% 的混合测试集。当网络仅由 BayarConv2D、Block1、Block2 组成时，

表1 消融实验结果

网络块设置							检测准确率
BayarConv2D	Block1	Block2	Block3	Block4	PAM	CAM	
√	√						92.48%
	√	√					92.32%
	√		√				93.76%
	√			√			93.55%
√	√	√					93.21%
√	√	√	√				94.92%
√	√	√		√			94.79%
√	√	√	√	√			96.43%
√	√	√	√	√	√		96.87%
√	√	√	√	√		√	96.92%
√	√	√	√	√	√	√	97.84%

实验的准确率为93.21%，效果图如图5（a）所示；加入Block3模块后，准确率为94.92%，效果图如图5（b）所示；再加入Block4模块后，准确率为96.43%，效果图如图5（c）所示；再加入PAM模块和CAM模块后，准确率为97.84%，效果图如图5（d）所示。逐步加入所提出的模块，到分类边界越来越明显，分类性能逐步提升，准确率逐步提升，这证明了所提模块的有效性。

3.5 对比实验

在本节中，将所提出的网络模型与其他主流方法进行比较，其中包括传统方法^[11]和4种基于深度学习的方法^[14-17]。表2展示了在不同缩放因子情况下6种检测方法的准确度，其中最好的结果加粗表示。

在低缩放因子情况下，图像中Seam Carving痕迹很少，检测工作存在困难。通过表2可以发现，

与其他5种检测方法相比，本文提出的网络模型在低缩放因子为1%~9%情况下均达到了最好的效果，甚至在1%时达到了89.48%的检测准确率，这充分证明了本文网络模型的有效性。

在高缩放因子情况下，随着缩放因子的增大，图像会严重失真，通常能用肉眼判断，检测难度降低。与其他5种方法相比，本文提出的网络模型除了在缩放因子为40%时准确率略低于HPF-CNN^[15]，其他情况下均达到了最好的效果，这证明了本文网络模型的有效性。

3.6 计算复杂度分析

本节通过计算每个网络的可训练参数量和浮点运算次数（floating point operations, FLOPs），分析和比较所提出网络与4个神经网络的计算复杂度。参数量指的是模型中需要学习的参数数量，可以通过该层中的权重矩阵的大小加上偏置向量的大小来计算，其大小直接影响着模型的存储空间和计算量。FLOPs通常用来衡量模型的计算复杂度，即完成一次前向传播所需的浮点运算次数。如表3所示，与其他主流方法相比，本文所提出的网络在准确性方面表现出最佳性能，同时具有较低的计算复杂度。具体来说，本文模型由0.43 M个可训练参数组成，训练一次需要进行0.67 M个FLOPs。图6显示了检测准确率与计算复杂度之间的散点图。

3.7 鲁棒性实验

图像伪造者在篡改图像后，往往会采用一系列后处理操作隐藏篡改痕迹，通常选择将图片压缩成

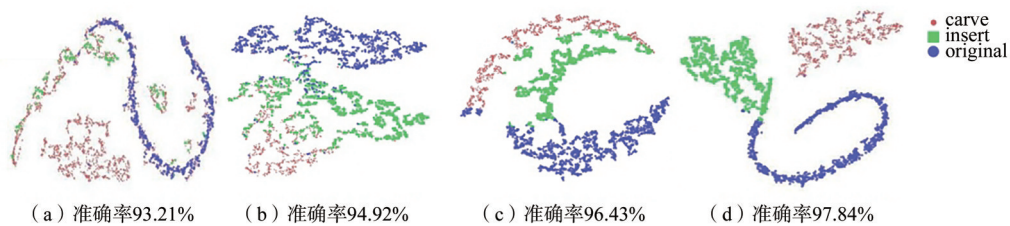


图5 部分消融实验的可视化效果

表2 对比实验结果

方法	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	20%	30%	40%	50%
BACM-SVM ^[11]	61.17%	64.38%	66.29%	67.89%	70.53%	72.09%	75.27%	76.59%	78.46%	80.33%	86.47%	90.26%	94.19%	97.88%
LBP-CNN ^[14]	61.37%	62.96%	64.46%	67.12%	72.49%	75.67%	78.34%	80.44%	81.37%	83.19%	87.39%	93.58%	97.82%	98.90%
HPF-CNN ^[15]	70.16%	74.85%	79.23%	84.56%	90.12%	91.40%	92.53%	93.71%	94.89%	95.63%	97.82%	98.69%	99.74%	99.86%
SRNet ^[16]	63.16%	65.89%	68.27%	71.54%	73.75%	76.40%	79.16%	82.90%	85.25%	87.60%	93.54%	96.77%	98.16%	98.92%
ILFNet ^[17]	66.47%	69.54%	72.16%	74.89%	77.36%	81.19%	83.60%	85.37%	88.29%	90.46%	95.79%	98.34%	99.57%	99.74%
Ours	89.48%	92.53%	94.20%	95.36%	95.81%	96.63%	97.23%	97.35%	97.94%	98.21%	98.85%	99.48%	99.63%	99.89%

表3 网络计算复杂度分析结果

方法	检测准确率	参数量	FLOPs
LBP-CNN ^[14]	84.19%	2.67 M	6.83 M
HPF-CNN ^[15]	94.63%	1.12 M	2.25 M
SRNet ^[16]	87.60%	4.63 M	9.37 M
ILFNet ^[17]	92.46%	0.36 M	0.74 M
Ours	97.21%	0.43 M	0.86 M

一定质量因子的JPEG格式。因此,验证模型对于JPEG压缩后处理的鲁棒性至关重要。本文使用Matlab软件将上述得到的图像数据集按照质量因子(quality factor, QF)为90、80、70、60和50进行JPEG压缩处理后,使用训练好的模型进行测试,不同缩放因子情况下图像检测准确率如表4所示。尽管在缩放因子较低时且经过JPEG进一步压缩处理后,图像会丢失部分Seam Carving操作痕迹,但是除了缩放因子为1%和2%的情况检测准确率较低,其他均能达到80%以上,而在高缩放因子(缩放因子大于10%)情况下,即使质量因子低至50,检测准确率仍在90%以上,这充分证明所提网络模型对于JPEG压缩后处理具有较强的鲁棒性。

3.8 泛化性实验

模型的泛化能力也是评价一个模型好坏的重要

标准。为了验证本文模型的泛化性,同时在UCID数据集^[29]和实拍图像数据集上进行了测试。

(1) UCID数据集的泛化性实验

UCID数据集包含1 338张彩色图像,空间分辨率为512×384,采用TIFF格式。实验同样在UCID数据集上进行1%~50%的垂直消缝操作和垂直缝插入操作,图像使用质量因子为100的JPEG压缩保存。UCID数据集的检测准确率如表5所示。可以看出,当缩放因子为1%时,检测准确率为85.47%,在低缩放因子情况下仍然具有较高的准确率;在缩放因子为50%时,准确率高达99.81%,与原始数据集BOSSbase 1.01具有相似的结果。由此可见,本文提出的网络模型具有良好的泛化能力。

(2) 实拍图像数据集的泛化性实验

实拍图像数据集包括100张彩色图片,其中风景图34张、建筑图47张、人像图19张,空间分辨率为512×384,采用JPG格式。图7展示了实拍图像数据集部分图像。同样在实拍图像数据集上进行1%~50%的垂直消缝操作和垂直缝插入操作,并且图像使用质量因子为100的JPEG压缩保存。实拍图像数据集的检测准确率如表6所示。当缩放因子为1%时,检测准确率为83.91%;当缩放因子为50%时,准确率为99.67%。

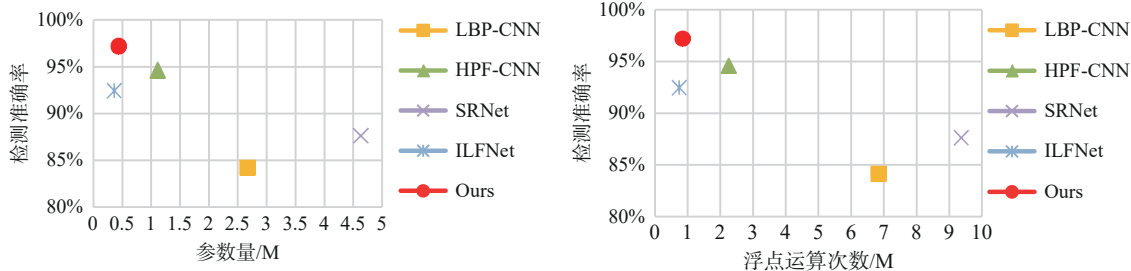


图6 网络的检测准确率与计算复杂度之间的散点图

表4 不同缩放因子情况下图像检测准确率

质量因子	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	20%	30%	40%	50%
QF90	88.51%	89.63%	93.46%	95.12%	94.36%	94.72%	95.39%	96.28%	96.77%	96.89%	97.47%	98.54%	99.03%	99.42%
QF80	77.38%	86.47%	88.15%	91.63%	92.47%	92.95%	93.12%	93.89%	94.16%	95.36%	96.39%	97.76%	98.40%	98.73%
QF70	71.36%	83.26%	83.62%	87.42%	89.14%	90.13%	91.29%	92.46%	93.05%	93.45%	94.52%	95.33%	96.19%	96.82%
QF60	66.94%	76.39%	81.89%	84.94%	87.66%	88.32%	88.93%	89.47%	90.18%	91.71%	92.66%	93.21%	94.54%	94.85%
QF50	66.55%	71.42%	80.34%	82.13%	83.29%	86.61%	87.46%	88.51%	89.72%	90.12%	91.23%	92.16%	93.42%	93.98%

表5 UCID数据集的检测准确率

数据集	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	20%	30%	40%	50%
UCID	85.47%	87.63%	90.25%	91.49%	92.66%	93.29%	93.84%	94.53%	95.14%	95.69%	96.72%	98.53%	99.34%	99.81%

表6 实拍图像数据集的检测准确率

数据集	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	20%	30%	40%	50%
实拍图像数据集	83.91%	86.24%	88.59%	89.32%	91.18%	91.83%	92.55%	93.36%	93.92%	94.70%	96.16%	97.83%	98.91%	99.67%



图7 实拍图像数据集部分图像

4 结论

本文提出了一种基于混合注意力机制的 Seam Carving 篡改检测网络模型。首先利用 BayarConv2D 对图像进行预处理, 并利用 10 个残差块进行特征学习, 利用混合注意力机制同时提取相邻位置和通道之间的特征, 最终将特征输入全连接层中进行分类。实验证明: 本文提出的网络模型具有很高的检测精度, 在缩放因子为 1% 时能到达 89.48% 的准确率, 远远高于目前主流方法的准确率; 在保证高准确率的情况下, 本文提出的网络模型具有较低的计算复杂度, 并且具有较好的鲁棒性, 能够抵抗 JPEG 压缩攻击, 除了缩放因子为 1% 和 2% 的情况, 其他情况的检测精度均能达到 80% 以上。未来的工作将专注于增强模型对其他后处理攻击的鲁棒性, 解决 Seam Carving 目标移除操作的篡改区域定位问题。

参考文献:

- [1] 吴晶辉, 严彩萍, 李红, 等. 边缘引导的双注意力图像拼接检测网络[J]. 中国图象图形学报, 2024, 29(2): 430-443.
- [2] WU J H, YAN C P, LI H, et al. BDA-Net: boundary-guided dual attention network for image splicing detection[J]. Journal of Image and Graphics, 2024, 29(2): 430-443.
- [3] VIJAYALAKSHMI K N V S K, SASIKALA J, SHANMUGANATHAN C. Copy-paste forgery detection using deep learning with error level analysis[J]. Multimedia Tools and Applications, 2024, 83(2): 3425-3449.
- [4] 李莉, 曾伟良, 黄永慧, 等. 基于半监督学习的人脸识别反欺骗方法研究[J]. 智能科学与技术学报, 2021, 3(3): 370-380.
- [5] LI L, ZENG W L, HUANG Y H, et al. Research on anti-spoofing method of face recognition based on semi-supervised learning[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(3): 370-380.
- [6] GHOSH A, REDDY N D, MERTZ C, et al. Learned two-plane perspective prior based image resampling for efficient object detection [C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 13364-13373.
- [7] 赵俊杰, 王金伟, 吴俊凤. 基于多质量因子压缩误差的对抗样本攻击方法识别[J]. 中国图象图形学报, 2023, 28(3): 850-863.
- [8] ZHAO J J, WANG J W, WU J F. Adversarial attack method identification model based on multi-factor compression error[J]. Journal of Image and Graphics, 2023, 28(3): 850-863.
- [9] AGARWAL S, JUNG K H. Median filtering detection using optimal multi-direction threshold on higher-order difference pixels[J]. Multimedia Tools and Applications, 2023, 82(20): 30875-30893.
- [10] PENG Y F, LIU X X, WANG G, et al. Irregular shape image retargeting based on seam carving[C]//Proceedings of the 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI). Piscataway: IEEE Press, 2023: 1845-1848.
- [11] LI L, XUE J R, TIAN Z Q, et al. Moment feature based forensic detection of resampled digital images[C]//Proceedings of the 21st ACM International Conference on Multimedia. New York: ACM, 2013: 569-572.
- [12] SHAMIR A, AVIDAN S. Seam carving for media retargeting[J]. Communications of the ACM, 2009, 52(1): 77-85.
- [13] SARKAR A, NATARAJ L, MANJUNATH B S. Detection of seam carving and localization of seam insertions in digital images[C]//Proceedings of the 11th ACM Workshop on Multimedia and Security. New York: ACM, 2009: 107-116.
- [14] WATTANACHOTE K, SHIH T K, CHANG W L, et al. Tamper detection of JPEG image due to seam modifications[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(12): 2477-2491.
- [15] YIN T, YANG G B, LI L D, et al. Detecting seam carving based image resizing using local binary patterns[J]. Computers & Security, 2015, 55: 130-141.
- [16] KARAPINAR SENTURK Z, AKGUN D. A machine learning method for detecting the trace of seam carving[J]. Elektronika Ir Elektrotehnika, 2021, 27(5): 59-66.
- [17] FERNANDODA SILVA CIESLAK L, AUGUSTOPONTARA DA COSTA K, PAULO PAPA J. Seam carving detection using convolutional neural networks[C]//Proceedings of the 2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI). Piscataway: IEEE Press, 2018: 195-200.
- [18] YE J Y, SHI Y X, XU G S, et al. A convolutional neural network based seam carving detection scheme for uncompressed digital images[C]//Proceedings of International Workshop on Digital Watermarking. Cham: Springer, 2019: 3-13.
- [19] BOROUMAND M, CHEN M, FRIDRICH J. Deep residual network for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1181-1193.
- [20] NAM S H, AHN W, YU I J, et al. Deep convolutional neural network for identifying seam-carving forgery[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(8): 3308-3326.
- [21] NATARAJ L, GUDAVALLI C, MANHAR MOHAMMED T, et al. Seam carving detection and localization using two-stage deep neural networks[C]//Proceedings of Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication. Singapore: Springer, 2021: 381-394.
- [22] MOREIRA T P, SANTANA M C S, PASSOS L A, et al. An end-to-end approach for seam carving detection using deep neural networks[C]//

- Proceedings of Iberian Conference on Pattern Recognition and Image Analysis. Cham: Springer, 2022: 447-457.
- [20] XIA M, CHEN J Y, YANG G B, et al. Robust detection of seam carving with low ratio via pixel adjacency subtraction and CNN-based transfer learning[J]. Journal of Information Security and Applications, 2023, 75: 103522.
- [21] BAYAR B, STAMM M C. Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2691-2706.
- [22] ZHANG R Y, NI J Q. A dense U-net with cross-layer intersection for detection and localization of image forgery[C]//Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2020: 2982-2986.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [24] ZHANG D Y, ZHU W J, DING X L, et al. SRTNet: a spatial and residual based two-stream neural network for deepfakes detection[J]. Multimedia Tools and Applications, 2023, 82(10): 14859-14877.
- [25] CELEBI N H, HSU T L, LIU Q Z. A comparison study to detect seam carving forgery in JPEG images with deep learning models[J]. Journal of Surveillance, Security and Safety, 2022, 3(3): 88-100.
- [26] GUDAVALLI C, ROSTEN E, NATARAJ L, et al. SeeTheSeams: localized detection of seam carving based image forgery in satellite imagery [C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2022: 1-11.
- [27] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2018: 3-19.
- [28] BAS P, FILLER T, PEVNÝ T. "Break our steganographic system": the

ins and outs of organizing BOSS[C]//Proceedings of International Workshop on Information Hiding. Heidelberg: Springer, 2011: 59-70.

- [29] SCHAEFER G, STICH M. UCID: an uncompressed color image database[C]//Proceedings of Storage and Retrieval Methods and Applications for Multimedia 2004. [S.l.:s.n.], 2003: 472-480.

[作者简介]



赵洁 (1984-), 男, 博士, 天津城建大学计算机与信息工程学院副教授, 主要研究方向为图像处理、计算机视觉、机器学习。



常皓婵 (2000-), 女, 天津城建大学计算机与信息工程学院硕士生, 主要研究方向为图像取证、机器学习。



武斌 (1966-), 男, 天津城建大学计算机与信息工程学院教授, 主要研究方向为计算机视觉、模式识别。