

基于双分支网络的图像修复取证方法

章登勇^{1,2}, 文凰^{1,2}, 李峰^{1,2}, 曹鹏^{1,2}, 向凌云^{1,2}, 杨高波³, 丁湘陵⁴

(1. 长沙理工大学湖南省综合交通运输大数据智能处理重点实验室, 湖南 长沙 410114;
2. 长沙理工大学计算机与通信工程学院, 湖南 长沙 410114; 3. 湖南大学信息科学与工程学院, 湖南 长沙 410082;
4. 湖南科技大学计算机科学与工程学院, 湖南 湘潭 411004)

摘要: 图像修复是一项利用图像已知区域的信息来修复图像中缺失或损坏区域的技术。人们借助以此为基础的图像编辑软件无须任何专业基础就可以轻松地编辑和修改数字图像内容, 一旦图像修复技术被用于恶意移除图像的内容, 会给真实的图像带来信任危机。目前图像修复取证的研究只能有效地检测某一种类型的图像修复。针对这一问题, 提出了一种基于双分支网络的图像修复被动取证方法。双分支中的高通滤波卷积网络先使用一组高通滤波器来削弱图像中的低频分量, 然后使用4个残差块提取特征, 再进行两次4倍上采样的转置卷积对特征图进行放大, 此后使用一个 5×5 的卷积来减弱转置卷积带来的棋盘伪影, 生成图像高频分量上的鉴别特征图。双分支中的双注意力特征融合分支先使用预处理模块为图像增添局部二值模式特征图。然后使用双注意力卷积块自适应地集成图像局部特征和全局依赖, 捕获图像修复区域和原始区域在内容及纹理上的差异, 再对双注意力卷积块提取的特征进行融合。最后对特征图进行相同的上采样, 生成图像内容和纹理上的鉴别特征图。实验结果表明该方法在检测移除对象的修复区域上, 针对样本块修复方法上检测的 $F1$ 分数较排名第二的方法提高了2.05%, 交并比上提高了3.53%; 针对深度学习修复方法上检测的 $F1$ 分数较排名第二的方法提高了1.06%, 交并比提高了1.22%。对结果进行可视化可以看出, 在检测修复区域上能够准确地定位移除对象的边缘。

关键词: 图像取证; 图像修复检测; 深度学习; 注意力机制

中图分类号: TP393

文献标识码: A

DOI: 10.11959/j.issn.2096-109x.2022084

Image inpainting forensics method based on dual branch network

ZHANG Dengyong^{1,2}, WEN Huang^{1,2}, LI Feng^{1,2}, CAO Peng^{1,2},

收稿日期: 2022-03-27; 修回日期: 2022-08-17

通信作者: 李峰, lif@csust.edu.cn

基金项目: 国家自然科学基金(62172059, 61972057, 62072055); 湖南省自然科学基金(2020JJ4626, 2020JJ4029); 湖南省教育厅优秀青年项目(19B004)

Foundation Items: The National Natural Science Foundation of China (62172059, 61972057, 62072055), Natural Science Foundation of Hunan Province (2020JJ4626, 2020JJ4029), Scientific Research Fund of Hunan Provincial Education Department of China (19B004)

引用格式: 章登勇, 文凰, 李峰, 等. 基于双分支网络的图像修复取证方法[J]. 网络与信息安全学报, 2022, 8(6): 110-122.

Citation Format: ZHANG D Y, WEN H, LI F, et al. Image in painting forensics method based on dual branch network[J]. Chinese Journal of Network and Information Security, 2022, 8(6): 110-122.

XIANG Lingyun^{1,2}, YANG Gaobo³, DING Xiangling⁴

1. Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation,
Changsha University of Science and Technology, Changsha 410114, China

2. School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, 410114, China

3. School of Information Science and Engineering, Hunan University, Changsha 410082, China

4. School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411004, China

Abstract: Image inpainting is a technique that uses information from known areas of an image to repair missing or damaged areas of the image. Image editing software based on it has made it easy to edit and modify the content of digital images without any specialized foundation. When image inpainting techniques are used to maliciously remove the content of an image, it will cause confidence crisis on the real image. Current researches in image inpainting forensics can only effectively detect a certain type of image inpainting. To address this problem, a passive forensic method for image inpainting was proposed, which is based on a two-branch network. The high-pass filtered convolutional network in the dual branch first used a set of high-pass filters to attenuate the low-frequency components in the image. Then features were extracted using four residual blocks, and two transposed convolutions were performed with 4x up-sampling to zoom in on the feature map. And thereafter a 5×5 convolution was used to attenuate the tessellation artifacts from the transposed convolutions to generate a discriminative feature map on the high-frequency components of the image. The dual-attention feature fusion branch in the dual branch first added a local binary pattern feature map to the image using a preprocessing block. Then the dual-attention convolution block was used to adaptively integrate the image's local features and global dependencies to capture the differences in content and texture between the inpainted and pristine regions of the image. Additionally, the features extracted from the dual-attention convolution block were fused, and the feature maps were up-sampled identically to generate the discriminative image content and texture on the feature maps. The extensive experimental results show the proposed method improved the *F1* score by 2.05% and the Intersection over Union(IoU) by 3.53% for the exemplar-based method and by 1.06% and 1.22% for the deep-learning-based method in detecting the inpainted region of the removed object. Visualization of the results shows that the edges of the removed objects can be accurately located on the detected inpainted area.

Keywords: image forensics, image forgery detection, deep learning, attention mechanism

0 引言

随着移动互联网的飞速发展，多媒体应用软件成为人们生活中必不可少的应用。基于多媒体应用软件，如 Photoshop、美图秀秀、醒图等，人们无须任何专业基础就可以轻松编辑和修改数字图像的内容，并且修改后的图像十分逼真，无法通过肉眼检测出修改的痕迹。人们的学习、生活、娱乐都离不开数字图像，但伪造的图像频繁滥用，这种行为带来了许多负面的影响，验证图像是否真实完整的数字图像取证技术对社会显得越来越重要。对于一些典型的伪造图像的检测可以从图像中提取对应的特征来判断图像是否被篡改。例如，复制移动伪造检测从不同方面在图像中找到

相似的特征来定位被篡改的区域，拼接伪造检测在图像中寻找光照、数据统计以及相机响应函数等不一致的特征来定位被篡改的区域。在 20 世纪 80 年代，研究人员从古代文物、书画的修复工作中得到启发，提出了对数字图像进行修复的技术。图像修复技术是使用图像原始区域的信息来补全图像中缺失或损坏区域的过程。如果人们将这项技术用于图像中的对象去除，使用图像中已不包含对象的信息来修复去除的对象区域，图像的原意将会改变。利用这项技术恶意篡改图片内容，将会带来图像内容的信任危机，不利于数字图像技术的进一步发展。因此，研究一个高效的检测图像修复的取证方法在当前背景下是非常有意义的。

目前，图像修复技术分为两大类：基于传统

方法的图像修复技术和基于深度学习方法的图像修复技术。基于传统方法的图像修复技术中最常用的是基于扩散的^[1-3]和基于样本块的^[4-6]图像修复技术。基于扩散的图像修复技术利用缺失区域的边缘信息来确定扩散方向,确定方向后再将信息平滑地传播到缺失区域。这种技术存在一定的缺点,当其用于填充实景图像中的大块空白时,填充的纹理会变得模糊,肉眼即可看出修改区域。基于样本块的图像修复在图像背景简单时可以填充图像中的大面积缺失区域。对于基于样本块的图像修复技术,许多研究人员已经提出精度较高的检测方法,其中大多数方法在给定区域中搜索相似块来确定修复区域,这会消耗大量的时间。因此,一些研究人员尝试使用深度学习的方法来检测样本块图像修复,可以大大减少检测时间,提高整体检测效率。利用深度学习技术的图像修复^[7-9]甚至能够在复杂背景中填补大面积的缺失。针对深度学习的图像修复,研究人员采用深度学习来进行检测,该方面的研究仍然较少。目前缺少一个能够同时准确检测样本块图像修复和深度学习图像修复的方法。

为了解决上述问题,本文提出了一个双分支的神经网络架构,用于检测图像的修复区域,具体贡献如下。

1) 创建了样本块图像修复和深度学习图像修复的数据集以及符合实际应用情形的、对象移除的、采用样本块和深度学习修复的图像数据集。

2) 本文方法能够同时检测样本块图像修复和深度学习图像修复,在性能上优于一些先进的方法。

3) 本文方法不仅能够清晰地定位修复图像中移除对象的边界,还能准确检测出经过后处理的图像修复区域。在图像修复的实际应用中具有优异的性能。

1 相关工作

1.1 样本块图像修复取证

传统方法检测样本块图像修复已有不少成果。Wu 等^[10]首先提出了一种检测样本块修复的被动取证方法。该方法首先对图像中可疑区域的所有块的零连通性作为块的匹配特征;然后使用

升半梯形分布函数,确定块是否属于相似块;最后通过给定的阈值划分切割集将相似块分为原始块和修复块,聚集的块属于修复区域。此后,Bacchuwar 等^[11]提出在计算图像块的匹配度之前,先比较块的亮度分量和中值,跳过可疑区域中的一个或多个块,这样可以大大减少算法时间。上述方法需要给定图像的可疑区域,并且在可疑区域中搜索时会受到原始区域的内部相似区域(如天空、草地等)干扰,导致产生误报区域。针对相似区域,Chang 等^[12]提出使用相似向量删除可疑区域中具有纹理一致性的区域,然后使用多区域关系来判断图像的修复区域,在搜索相似块时,还提出了一种基于权重变换的两阶段搜索算法,加快相似度的计算。然而,基于权重变换的方法无法同时优化负载因子和搜索范围。Liang 等^[13]提出了基于中心像素映射的快速搜索方法来提高搜索速度。但是,伪造图像通常会经历后处理操作,图像的修复痕迹将被隐藏,因此该方法在检测图像修复区域上仍然存在一定的局限性。

Zhao 等^[14]提出通过计算压缩的修复图像中压缩次数的之间差异来检测修复区域。Liu 等^[15]在离散余弦变换(DCT, discrete cosine transformer)域中提取近十万个特征来定位图像修复区域,能够有效地检测重压缩下的图像修复伪造。但是,这两种方法只能检测经历了 JPEG 压缩后处理操作的修复图像,无法检测其他后处理操作。Zhang 等^[16]提出一种混合取证策略来检测经过多个后处理操作的修复伪造,但它无法定位图像的修复区域。

Zhu 等^[17]首次提出了一种基于卷积神经网络的方法来检测样本块图像修复篡改。卷积神经网络遵循编码器-解码器的网络结构构建。该方法可以在保持检测精度的同时大大减少检测时间。后来,Wang 等^[18]改进了语义分割中性能优秀的基于掩膜的区域卷积神经网络(Mask R-CNN, mask region-based convolutional neural network)检测图像修复篡改并识别修复区域。为了进一步利用所有尺度的特征信息,Wang 等^[19]提出在 Mask R-CNN 中将金字塔网络和反向连接相结合以提取更多特征。此外,Lu 等^[20]提出一种基于长短期记忆-卷积神经网络的图像修复检测方法,利

用卷积神经网络识别篡改图像中的可疑相似块，再使用长短期记忆网络识别图像中正常的纹理一致的区域，去除虚假可疑的相似块，降低误报率。

1.2 深度学习图像修复取证

随着深度学习的发展，研究人员提出了基于深度学习的图像取证方法来检测伪造的数字图像。起初，研究人员只是使用简单的卷积神经网络进行图像取证^[21,22]。随后，一些研究发现在卷积神经网络的输入中使用特殊的特征代替图像内容会获得更好的图像取证性能，如噪声残差^[23]、直方图特征^[24]、约束滤波器^[25]等。除了简单的卷积神经网络与输入特征的改变外，研究人员尝试使用长短期记忆架构在像素级别上定位篡改区域^[26-27]或者使用多分支卷积神经网络检测伪造区域^[28-29]，这给图像修复取证提供了很好的指导作用。

对于深度学习图像修复取证，由于修复区域和原始区域在感知上是一致的，而且不同于样本块图像修复，深度学习图像修复取证没有传统方法可以提取修复区域与原始区域的不一致性特征。Li等^[30]发现修复图像在高通滤波后的残留域中修复区域和原始区域中高频分量存在显著差异，利用这一特征，提出了高通滤波图像残差的全卷积网络来定位深度学习图像修复篡改的区域。Wang等^[31]提出基于更快的区域卷积神经网络（Faster R-CNN, faster region-based convolutional neural network）的语义分割网络来检测深度

学习图像修复的方法，该全卷积网络模型可以捕获修复区域和真实区域的差异特征，在检测深度学习图像修复上表现良好的性能。Li等^[32]发现噪声残留域中修复区域和原始区域具有明显的差异，基于噪声域提出了一种用于通用深度学习修复检测的新框架，可以识别由不同深度学习修复方法修复的区域。Wu等^[33]提出了一个深度学习图像修复检测网络，网络由增强块、提取块和决策块组成，可以准确检测和定位修复篡改。

尽管在深度学习图像修复取证上，已经有不少研究成果，但是目前缺少能够同时检测样本块图像修复和深度学习图像修复的方法。

2 双分支网络检测模型

本文提出了一个双分支网络检测模型，它能够准确检测图像修复区域。该模型的一个分支是高通滤波卷积网络（HPCN, high-pass filter convolutional network），参考文献[30]中的网络结构；另一个分支是双注意力特征融合网络。高通滤波卷积网络分支主要用于捕获图像修复区域与原始区域之间的高频分量差异，而双注意力特征融合（DAFF, dual-attention feature fusion）分支主要负责捕获图像修复区域与原始区域之间内容及纹理上的差异。整体的网络结构模型如图1所示。

2.1 高通滤波卷积网络分支

图像中的高频分量对应图像变化剧烈的部分，即图像的边缘、噪声或者细节部分；而低频分量则与之相反，对应变化度较小的部分。高通

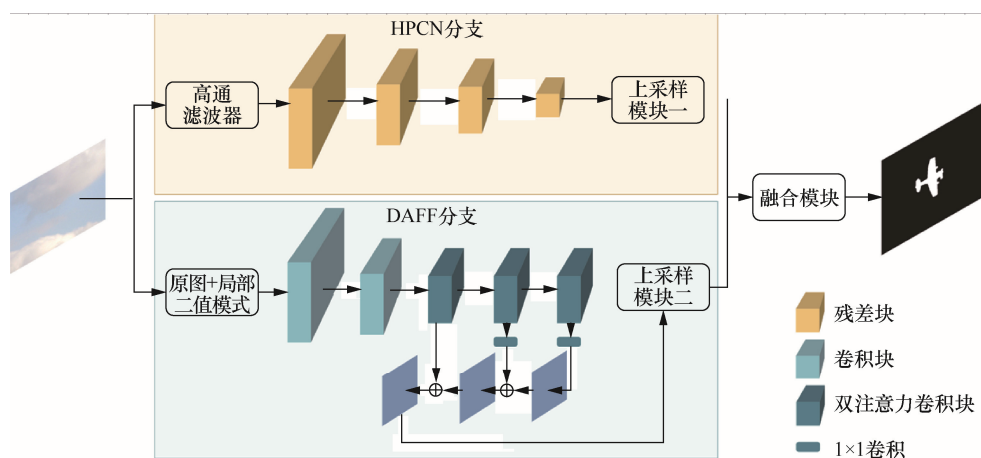


图1 网络结构
Figure 1 Network structure

滤波的作用是削弱低频分量并保持高频分量相对不变。在文献[30]中发现图像的修复区域相对于原始区域包含更少的高频分量,因此 HPCN 分支使用一组高通滤波器来削弱图像中的低频分量。高通滤波器的内核使用 3 个一阶高通滤波器初始化,不同于一般的卷积,这 3 个滤波器分别作用于 RGB 图像的每个通道,将产生的 9 个特征图作为后续输入。

残差网络^[34]可以解决深度网络的退化问题,因其性能优越常被用于图像分类和目标检测等许多计算机视觉应用。为了准确地从特征图中捕获出修复区域和原始区域中高频分量的差异,HPCN 分支使用 4 个残差块来提取特征,其参数配置如表 1 所示。每个残差块都由两个与 ResNet50 相同的瓶颈单元组成。瓶颈单元先用核大小为 1 的卷积对特征图进行降维,再用核大小为 3 的卷积提取特征,最后用核大小为 1 的卷积增加特征图的维数,参数量得到大幅减少。瓶颈单元中的恒等映射,用来将浅层特征图复制给深层高语义的特征图结合,通过残差块 HPCN 分支获得 1 024 个特征图,它们的分辨率是输入图像的 1/16。

HPCN 分支判断图像中每一个像素是属于原始区域还是修复区域,先使用以双线性插值为核的转置卷积扩大特征图的空间分辨率。为了避免计算量过大,HPCN 分支采用了两阶段策略,执行两次 4 倍上采样的转置卷积对特征图进行放大,将卷积核的大小从 32×32 减少到 8×8 。此后对特征图进行一次步长为 1 的 5×5 卷积来减弱转

置卷积带来的棋盘伪影,生成图像高频分量上的各个特征的加权和。

表 1 HPCN 分支中残差块的参数配置
Table 1 Parameter configuration of residual block in HPCN branch

残差块	瓶颈单元	瓶颈深度	输出深度	步长
#1	#1	32	128	1
	#2	32	128	2
#2	#1	64	256	1
	#2	64	256	2
#3	#1	128	512	1
	#2	128	512	2
#4	#1	256	1 024	1
	#2	256	1 024	2

2.2 双注意力特征融合分支

基于样本块的图像修复区别于深度学习的图像修复的一个显著特点是:基于样本块的图像修复会在图像全局搜索最相似的块进行填充。为了捕获相似性的块,DAFF 在预处理模块为图像增添了局部二值模式特征图。局部二值模式是一种用于描述图像局部纹理特征的算子,在图像取证中提取特征有一定的作用^[35],它表示某一像素点与它的 8 邻域构成的纹理特征,其计算公式为

$$\text{LBP}(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p - i_c) \quad (1)$$

其中 (x_c, y_c) 表示中心像素, i_c 是中心像素的灰度值, i_p 是中心像素相邻像素的灰度值, $s(\cdot)$ 是符号函数,为

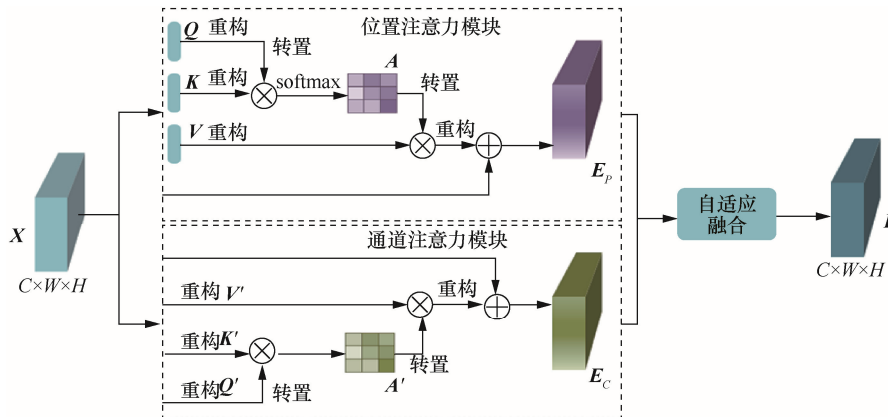


图 2 双注意力机制结构

Figure 2 Structure diagram of dual attention mechanism

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{其他} \end{cases} \quad (2)$$

当基于样本块的图像修复将最匹配的块填充到缺失区域时, 纹理特征将在缺失区域内部传播, 因此修复区域的局部二值模式值和原始区域中对应位置的局部二值模式值相同。为图像添加局部二值模式特征图可以帮助捕获图像修复区域与原始区域之间纹理上的差异。

文献[28]对几种主流的主干网络进行对比, 发现可视几何组 (VGG, visual geometry group) 在篡改区域的特征提取上有着更高的适用性。因此, DAFF 分支中基于 VGG-16 构建了特征提取模块, 首先将最后的 3 个全连接层去掉, 然后将卷积层划分为 5 个部分, 前 2 个部分为卷积块, 后 3 个部分的卷积块添加双重注意力机制^[36], 称为双注意力卷积块。在每个卷积块最后都使用了最大池化来减少网络的参数, 在双注意力卷积块中, 只有第一个双注意力卷积块使用最大池化。池化层虽然可以通过减少网络的参数来实现提高训练效率, 但是同样有降低特征图的空间分辨率、丢弃特征图部分信息等不可忽视的缺点。因此, 在最后两个双注意力卷积块中去掉了池化层, 并且在第二个双注意力卷积块中使用了空洞卷积。空洞卷积可以丰富深层特征的空间信息, 并生成高分辨率的特征图。DAFF 分支中特征提取模块的参数配置如表 2 所示。

双重注意力机制分为位置注意力模块和通道注意力模块, 采用的是自注意力机制。自注意力机制是从自然语言处理领域借鉴的思想, 因此仍然采用 Query、Key 和 Value 的命名方式。位置注意力特征图 E_p 中每个位置的值是通过原始特征图 X 每个位置的值进行加权求和, 再与原始特征相加得到的。首先, 将特征图 X 分别通过 3 个卷积, 输出特征图 Q 、 K 、 V , 其中 $\{Q, K, V\} \in \mathbb{R}^{C \times W \times H}$, 为了求出位置注意力, 将它们重构为 $\{Q, K, V\} \in \mathbb{R}^{C \times N}$, 其中 $N = H \times W$; 然后对 Q 和 K 进行相似度计算得到权重, 这里使用点积作为相似度函数; 最后使用 softmax 函数对权重进行归一化, 得到位置注意力图 $A \in \mathbb{R}^{N \times N}$, 其计算公式为

$$A = \text{softmax}(Q^T K) \quad (3)$$

模块将 V 和转置后的 A 相乘, 并将结果重构为 $\mathbb{R}^{C \times W \times H}$ 。位置注意力特征图 E_p 可以计算为

$$E_p = \alpha(\text{reshape}(VA^T)) + X \quad (4)$$

其中, α 为位置注意力特征图的权重, 初始值为 0, 会随着网络的训练而调整。

表 2 DAFF 分支中特征提取模块的参数配置
Table 2 Parameter configuration of the feature extraction module in DAFF branch

块	类型	配置
卷积块#1	卷积#1	核个数: 64, 大小: 3×3, 步长: 1
	卷积#2	核个数: 64, 大小: 3×3, 步长: 1
	最大池化	窗口大小: 2×2, 步长: 2
卷积块#2	卷积#1	核个数: 128, 大小: 3×3, 步长: 1
	卷积#2	核个数: 128, 大小: 3×3, 步长: 1
	最大池化	窗口大小: 2×2, 步长: 2
双注意力卷积块#1	卷积#1	核个数: 256, 大小: 3×3, 步长: 1
	卷积#2	核个数: 256, 大小: 3×3, 步长: 1
	卷积#3	核个数: 256, 大小: 3×3, 步长: 1
	最大池化	窗口大小: 2×2, 步长: 2
双重注意力 —		
双注意力卷积块#2	卷积#1	核个数: 512, 大小: 3×3, 步长: 1, 空洞卷积率: 2
	卷积#2	核个数: 512, 大小: 3×3, 步长: 1, 空洞卷积率: 2
	卷积#3	核个数: 512, 大小: 3×3, 步长: 1, 空洞卷积率: 2
双重注意力 —		
双注意力卷积块#3	卷积#1	核个数: 512, 大小: 3×3, 步长: 1
	卷积#2	核个数: 512, 大小: 3×3, 步长: 1
	卷积#3	核个数: 512, 大小: 3×3, 步长: 1
双重注意力 —		

通道注意力特征图 E_c 不同于位置注意力, 它是对每个通道进行加权求和, 再与原始特征图 X 相加。首先对特征图进行重构得 $\{Q', K', V'\} \in \mathbb{R}^{C \times N}$, 其中 $N = H \times W$; 然后对 Q' 和 K' 进行点积计算得到权重; 最后使用 softmax 函数对权重进行归一化, 通道注意力相似度计算是对通道进行计算, 所以得到的通道注意力图为 $A' \in \mathbb{R}^{C \times C}$, 其计算公式为

$$A' = \text{softmax}(K'Q'^T) \quad (5)$$

模块将 V' 和转置后的 A' 相乘, 并将结果重构为 $\mathbb{R}^{C \times W \times H}$ 。通道注意特征图 E_C 可以计算为

$$E_C = \beta(\text{reshape}(V'A'^T)) + X \quad (6)$$

其中, β 为通道注意力特征图的权重, 初始值为 0, 会随着网络的训练而调整。

为位置注意力特征图 E_p 和通道注意特征图 E_C 添加权重, 以便两个注意力特征图能够通过训练自适应调整各自的重要性。双重注意力模块的输出 E 定义为

$$E = \gamma E_p + \theta E_C \quad (7)$$

其中, γ 和 θ 是位置注意模块和通道注意模块的权重。权重由标准正态分布初始化并通过网络训练进行自适应。

DAFF 分支使用 3 个双注意力卷积块来提取特征, 将第 k 个双注意力卷积块得到的特征图命名为 $E^k \in \mathbb{R}^{C \times H \times W}$, 其中 $k \in \{1, 2, 3\}$ 。根据表 2 给出的参数配置, 第一个双注意力卷积块特征图 E^1 的通道数为 256, 其余两个双注意力卷积块特征图 E^2 和 E^3 的通道数为 512, 它们的大小均为原始图像的 1/8。为了融合 3 个双注意力卷积块的特征, 首先使用 1×1 的卷积将 E^2 和 E^3 的通道数转换为 256, 然后将处理后的 E^2 和 E^3 相加, 使用一个 3×3 的卷积消除特征图相加产生的混叠效应后再与 E^1 相加。将融合后的特征图先执行一次 2 倍双线性插值的上采样, 再执行一次 4 倍双线性插值的上采样。因为同 HPCN 一样使用转置卷积进行上采样, 所以也使用步长为 1 的 5×5 卷积来减弱转置卷积带来的棋盘伪影, 产生 DAFF 分支的输出。

2.3 融合模块

当两个分支捕获了高频分量、内容及纹理上修复区域和原始区域之间的差异后, 为更好地融合特征并进行最终的预测, 将两个分支的输出送入如图 3 所示的 Inception 模块进行特征融合。Inception 是 GoogleNet 中的模块^[25], 它使网络变宽, 既能减少计算开销和内存消耗又能丰富特征。融合模块的具体步骤如下。

- 1) 连接 HPCN 分支和 DAFF 分支的输出。
- 2) 使用 3 个不同大小的过滤器进行特征提取, 核大小为 1×1 、 3×3 、 5×5 。

- 3) 将特征图连接起来作为输出。

对融合模块的特征图再进行 softmax 分类, 以获得图像修复像素级别的定位图。

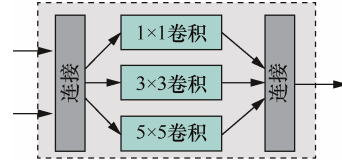


图 3 融合模块的结构
Figure 3 The illustration of the fusion module

2.4 损失函数

图像修复区域的定位是二分类问题, 其像素级的预测结果只有两种情况: 原始区域和修复区域。将修复区域视为正样本, 由于实验数据集的修复区域是随机生成的, 因此使用了 3 种常用的损失函数进行实验。实验结果表明, 标准交叉熵损失比焦点损失和加权交叉熵损失具有更好的性能。原因可能是数据集中随机生成的修复区域与原始区域的面积大致相同, 因此使用标准交叉熵损失函数作为模型的损失函数。训练数据 (X, Z) 的标准交叉熵损失为

$$L = \frac{-1}{W \cdot H} \sum_{j=1}^W \sum_{i=1}^H Z_{ij} \log \hat{Y}_{ij} + (1 - Z_{ij}) \log(1 - \hat{Y}_{ij}) \quad (8)$$

其中, $\hat{Y}_{ij}$ 表示预测概率矩阵 $\hat{Y}$ 中的第 i 行第 j 列元素, Z_{ij} 是标签矩阵中与之对应的元素。

3 实验结果与分析

使用 PASCAL VOC^[37] 创建了基于深度学习和基于样本块的普通图像修复数据集和对象移除图像修复数据集, 在数据集上训练测试了本文提出的方法, 并与图像修复检测的文献[17]、文献[30]、文献[33]中评估的性能指标 $F1$ 分数和交并比进行了比较。

3.1 实验数据集

目前暂无用于图像修复取证的公开数据集, 本文使用 PASCAL VOC 创建了样本块图像修复和深度学习的图像修复数据集。PASCAL VOC 包含 20 种特定对象的自然图像, 并提供有对象的像素级的分割图, 为创建对象移除的图像修复数据集提供了便利。首先将 PASCAL VOC 2007 和 PASCAL VOC 2012 训练集的图像混合去除重复

的图像,获得共 21 542 幅图像;然后采用 3 种方法创建图像的缺失区域,分别是:① 中心矩形区域;② 随机 1 到 5 个矩形区域;③ 随机一些线、圆、椭圆;最后分别采用基于样本块的图像修复方法^[4]和基于深度学习的图像修复方法^[9]对图像进行修复。图 4 展示了几个示例图像。将这一部分数据集 90% 的图像作为训练集,10% 的图像作为验证集。

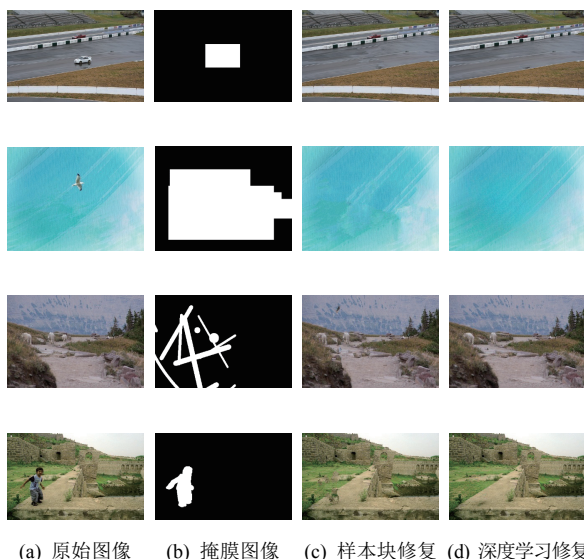


Figure 4 Several examples of the experimental dataset

将 PASCAL VOC 2007 和 PASCAL VOC 2012 测试集图像混合去除重复图像,同样采用训练集中的方法创建图像的缺失区域,并使用基于样本块的图像修复方法^[4]和基于深度学习的图像修复方法^[9]对图像进行修复,获得 8 915 幅修复图像作为测试集。另外,为了验证本文提出的方法是否符合实际情况,从 PASCAL VOC 2007 中的 300 幅对象像素级分割图中选取了一些对象(如人、猫、狗等)作为图像缺失区域进行修复,更准确地模拟现实生活中图像修复的用途。将这一部分数据集作为对象移除图像修复的测试集对网络效果进行验证。图像修复数据集的数量如表 3 所示。

3.2 实验设置

该实验的实验环境为 Ubuntu16.04, Nvidia GeForce RTX 2080ti GPU。实验中选择 Adam^[31]作为优化器,初始学习率设置为 0.000 1,此后的每个 epoch 学习率降低 80%。因为数据集图像大小不一致,所以将每一批次的大小设置为 1。

整个实验训练迭代 10 次,选择验证集结果中 $F1$ 分数值最高的模型作为最后模型。

表 3 图像修复数据集的数量
Table 3 Number table of image inpainting datasets

单位:幅

数据集	样本块图像修复数据集	深度学习图像修复数据集
训练集	19 388	19 388
验证集	2 154	2 154
随机缺失测试集	8 915	8 915
对象移除测试集	275	275

3.3 评估指标

本文主要使用两种逐像素分类指标来评估性能,分别是 $F1$ 分数和交并比。它们是测试集评估指标数据的平均值,计算如下。

$$r = \frac{TP}{TP+FN} \quad (9)$$

$$p = \frac{TP}{TP+FP} \quad (10)$$

$$F1 = \frac{2pr}{p+r} \quad (11)$$

其中,因为将修复区域视为正样本,TP 表示像素预测为修复区域,实际是修复区域的像素数量;FN 表示像素预测为原始区域,但实际是修复区域的像素数量;FP 表示像素预测为修复区域,实际是原始区域的像素数量。 r 表示召回率, p 表示精确率,交并比计算如下。

$$IoU = \frac{\text{intersection}}{\text{union}} \quad (12)$$

其中,intersection 表示图像预测区域和图像实际修复区域重合的像素数量,union 表示图像预测区域和图像实际修复区域相并的像素数量,IoU 表示交并比。

4.3 实验结果与分析

对于本文提出的方法,我们使用了焦点损失、加权交叉熵损失和标准交叉熵损失进行实验。在随机缺失的样本块图像修复和深度学习图像修复、对象移除的深度学习修复的数据集上进行测试,检测结果如表 4 所示。

加权交叉熵损失在修复区域和原始区域两类交叉熵损失前面添加了不同的权重。将像素错判

表 4 3 种损失在不同图像修复数据集上的检测结果
Table 4 Detection results of different losses on different image inpainting datasets

损失类型	随机缺失的样本块图像修复				随机缺失的深度学习图像修复			
	召回率	精确率	交并比	F1 分数值	召回率	精确率	交并比	F1 分数值
焦点损失	90.32%	91.07%	84.17%	89.98%	94.08%	96.62%	91.30%	95.29%
加权交叉熵损失	93.76%	88.82%	84.92%	90.46%	97.18%	93.99%	91.70%	95.53%
标准交叉熵损失	93.10%	91.79%	86.92%	91.72%	96.29%	97.38%	94.02%	96.78%

为原始区域的惩罚大于将像素错判为修复区域的惩罚。因此,使用加权交叉熵损失具有较高的召回率,但准确率较低。考虑到实际用途中修复区域可能小于未修复区域,在训练中使用焦点损失来增加修复区域的权重,并增加错误分类样本的权重。通过实验发现,焦点损失并不能有效提升性能。观察数据集发现,随机生成的修复区域与原始区域大小几乎相同,不存在需要焦点损失解决的正负样本不平衡问题。所以本文方法采用标准的交叉熵损失函数。

为了验证本文方法的有效性,对网络结构进行了消融实验,定量验证评估指标 $F1$ 分数值如表 5 所示。

消融实验各结构分解如下。

1) 1-Base: HPCN 分支,文献[18]中的高通全卷积网络。

2) 2-Base: 去除 DAFF 分支中双重注意力模块和特征融合模块的网络,即 DAFF 分支的基础网络结构。

3) 2-Base+DA: 在 DAFF 分支的基础网络结构上增加了双重注意力模块。

4) 2-Base+DA+FF: 在 DAFF 分支的基础网络结构上增加了双重注意力模块和特征融合模块。

5) 本文提出的网络: 带有 HPCN 分支和 DAFF 分支的双分支网络。

表 5 消融实验的检测结果
Table 5 Detection results of ablation experiments

网络	随机缺失的样本块 图像修复的 $F1$ 分数值	随机缺失深度学习 图像修复的 $F1$ 分数值
1-Base	89.568 8%	95.641 2%
2-Base	84.901 9%	95.684 4%
2-Base+DA	81.951 3%	95.937 8%
2-Base+DA+FF	86.352 6%	96.535 8%
本文提出的网络	91.724 6%	96.780 4%

分析表 5 中数据可以看出,虽然在增加了双注意力机制后,样本块图像修复检测的 $F1$ 分数值降低了,但是当导入双注意力模块和特征融合模块时, $F1$ 分数值提高。这可能是因为样本块图像修复的痕迹很大程度依赖于浅层的纹理,在浅层网络双重注意力机制提取了相似图像块的全局依赖,将其与深层的语义特征进行融合使得网络的检测结果更准确。这说明添加双重注意力机制和特征融合能够帮助网络有效地提取样本块图像修复区域与原始区域的相似块特征。深度学习图像修复检测的 $F1$ 分数值在导入双注意力模块和特征融合模块得到进一步提升。将 HPCN 分支和 DAFF 分支结合后,样本块图像修复和深度学习图像修复的 $F1$ 分数值得到了提升,这说明双分支网络不仅可以同时检测深度学习的图像修复区域和样本块的图像修复区域,还在检测图像修复区域的性能上得到了进一步的提升。

为了检测本文方法的性能优于现有先进的方法,将本文方法与文献[17]、文献[30]、文献[33]分别在样本块图像修复和深度学习图像修复数据集上重新训练,然后在随机缺失的样本块图像修复、对象移除的样本块图像修复、随机缺失的深度学习图像修复、对象移除的深度学习修复的数据集上进行测试,进行评估指标 $F1$ 分数和交并比的比较,结果如表 6 所示。

从表 6 中可以看出,基于双分支网络的图像修复取证算法在所有指标上都超过 3 种先进的方法。对于基于样本块技术的修复图像,无论是随机缺失的修复图像还是对象移除的修复图像,本文算法检测在 $F1$ 分数值提高了 2% 以上。特别是对于对象移除的修复图像,本文提出的算法比排名第二的算法在 $F1$ 分数值高 2.05%,在交并比上数值高 3.53%。对于基于深度学习技术的修复图

表6 不同模型在不同图像修复数据集上的检测结果
Table 6 Detection results of different models on different image inpainting datasets

方法	样本块图像修复				深度学习图像修复			
	随机缺失测试集		对象移除测试集		随机缺失测试集		对象移除测试集	
	F1 分数值	交并比	F1 分数值	交并比	F1 分数值	交并比	F1 分数值	交并比
文献[17]	89.69%	82.45%	87.53%	79.35%	91.60%	85.13%	90.64%	84.04%
文献[30]	89.56%	83.56%	91.29%	85.25%	95.64%	91.89%	95.57%	91.74%
文献[33]	62.37%	51.94%	31.36%	22.36%	96.58%	93.64%	95.11%	91.58%
本文	91.72%	86.91%	93.34%	88.78%	96.78%	94.02%	96.17%	92.80%

像,虽然本文方法较排名第二的算法提升不多,但是排名第二的算法在检测基于样本块技术的修复图像时,检测精度大大降低,甚至在对象移除的数据集上只有 31.36% 的 $F1$ 分数值。虽然该方法在深度学习图像修复中效果较好,但却在样本块修复的图像中忽略了样本块的相似性,导致其在样本块图像修复数据集上表现较差。这说明本文提出的算法不仅能够检测基于深度学习的图像修复,还能够检测基于样本块的图像修复,并取得了优越的性能。

为了更直观地感受本文方法的检测效果,本文从随机缺失的样本块图像修复、对象移除的样本块图像修复、随机缺失的深度学习图修复、对象移除的深度学习修复的数据集选取了多幅图像,展示了文献[17]、文献[30]、文献[33]和提出方法的预测掩模图,如图 5 所示。

从图 5 可以看出,在样本块图像修复中,文献[17]、文献[30]、文献[33]都无法准确定位修复区域边缘,这是因为它们忽略了样本块图像修复中图像块的相似性。在深度学习图像修复中,文献[30]和文献[33]都表现出优越的检测性能,但是本文提出方法能精确地分割出修复区域和原始区域。这说明本文提出的方法可以准确定位修复的移除对象区域的边界,表明了本文提出的方法不仅提高了图像修复检测的性能,而且在实际应用中表现出良好的性能。

在对图像进行篡改之后,为了掩盖篡改痕迹,通常会用 JPEG 压缩和高斯模糊进行后处理操作。在图像的传播中,也会经历 JPEG 压缩。为了验证本文提出方法的优越性,对数据集进行 JPEG 压缩和高斯模糊的后处理操作,然后选择检测样本块图像修复的文献和检测深度学习图像修复的

文献与本文方法进行比较。

文献[30]指出大多图像最低使用 75 的质量因子(QF, quality factor)进行压缩。质量因子越小,压缩率越大,图片质量越差,图像会失真。当质量因子低于 75 时,图像视觉质量过低,无法保证视觉感知上与原始图像一致。因此对于 JPEG 压缩后处理操作,分别使用 95 和 75 的质量因子对修复图像进行 JPEG 压缩处理。高斯模糊核大小代表邻域每个像素对当前处理像素的影响程度,高斯核的维数越高,图像的模糊程度越大。当高斯核大于 5×5 时,同样会使得图像过于模糊,无法保证图像的视觉感知。因此对于高斯模糊后处理操作,分别使用 3×3 和 5×5 的高斯核对修复图像进行高斯模糊操作。鲁棒性检测结果的 $F1$ 分数值如表 7 所示。

对于 JPEG 压缩的后处理操作,3 个模型的 $F1$ 分数值都随着质量因子的降低而降低,检测深度学习图像修复的文献[30]在检测样本块图像修复上,随着图像质量因子的降低 $F1$ 分数值大幅下跌;而检测样本块图像修复的文献[18]在检测深度学习图像修复上,随着图像质量因子的降低 $F1$ 分数值也出现了大幅下跌。但是本文提出的方法,在检测不同类型的样本修复上,随着图像质量因子的降低,其检测效果都优于该类型图像修复模型。对于高斯模糊的后处理操作,本文提出方法的检测结果也表现出优秀的性能,这说明本文提出方法的鲁棒性是最佳的。

5 结束语

本文提出了一个双分支的神经网络,它是一种新颖的图像修复盲取证方法。利用双分支



图 5 不同模型的实际检测结果
Figure 5 The actual detection results of different models

表 7 不同模型的鲁棒性检测结果的 F1 分数值
Table 7 F1 scores of robust detection results for different models

文献	样本块图像修复				深度学习图像修复			
	JPEG 压缩		高斯模糊		JPEG 压缩		高斯模糊	
	QF95	QF75	3×3 高斯核	5×5 高斯核	QF95	QF75	3×3 高斯核	5×5 高斯核
文献[17]	0.622 198	0.504 967	0.857 684	0.812 820	0.880 446	0.811 035	0.785 285	0.728 888
文献[18]	0.785 101	0.726 922	0.849 848	0.734 478	0.921 320	0.739 590	0.911 001	0.897 738
本文	0.791 670	0.773 700	0.885 564	0.868 461	0.937 303	0.880 019	0.932 124	0.924 232

同时捕获图像在高频分量、内容及纹理上修复区域和原始区域的差异, 并将结果进行多尺度融合检测图像的修复区域。可以有效地弥补当前缺少同时检测样本块图像修复和深度学习图像修复的方法, 并且它能够清晰地检测出修复图像中移除对象的区域边界, 适用于现实场景。大量的实验表明了该方法的有效性。虽然实验数据显示本文方法能够很好地识别两种图像修复算法检测的区域, 也能抵抗修复图像的后处理操作, 但是其在两种算法上的性能存在一定的差距, 在样本块图像修复上的性能还有待提升。未来将进一步提高图像修复检测的性能。

参考文献:

- [1] SRIDEVI G, SRINIVAS KUMAR S. Image inpainting based on fractional-order nonlinear diffusion for image reconstruction[J]. *Circuits, Systems, and Signal Processing*, 2019, 38(8): 3802-3817.
- [2] DOU L, QIAN Z, QIN C, et al. Anti-forensics of diffusion-based image inpainting[J]. *Journal of Electronic Imaging*, 2020, 29(4): 043026.
- [3] ZHANG Y, DING F, KWONG S, et al. Feature pyramid network for diffusion-based image inpainting detection[J]. *Information Sciences*, 2021, 572: 29-42.
- [4] 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): 1200-1212.
- [5] QUY TRAN B, VAN NGUYEN T, DUY TRAN D, et al. Accelerating exemplar-based image inpainting with GPU and CUDA[C]//2021 10th International Conference on Software and Computer Applications. 2021: 173-179.
- [6] XU T, HUANG T Z, DENG L J, et al. Exemplar-based image inpainting using adaptive two-stage structure-tensor based priority function and nonlocal filtering[J]. *Journal of Visual Communication and Image Representation*, 2022, 83: 103430.
- [7] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. *ACM Transactions on Graphics (ToG)*, 2017, 36(4): 1-14.
- [8] ZENG Y, FU J, CHAO H, et al. Learning pyramid-context encoder network for high-quality image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1486-1494.
- [9] SUVOROV R, LOGACHEVA E, MASHIKHIN A, et al. Resolution-robust large mask inpainting with fourier convolutions[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022: 2149-2159.
- [10] WU Q, SUN S-J, ZHU W, et al. Detection of digital doctoring in exemplar-based inpainted images[C]//2008 International Conference on Machine Learning and Cybernetics. 2008, 3: 1222-1226.
- [11] BACCHUWAR K S, RAMAKRISHNAN K. A jump patch-block match algorithm for multiple forgery detection[C]//2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s). 2013: 723-728.
- [12] 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): 57-71.
- [13] LIANG Z, YANG G, DING X, 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: 75-85.
- [14] ZHAO Y Q, LIAO M, SHIH F Y, et al. Tampered region detection of inpainting JPEG images[J]. *Optik*, 2013, 124(16): 2487-2492.
- [15] LIU Q, SUNG A H, ZHOU B, et al. Exposing inpainting forgery in jpeg images under recompression attacks[C]//2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). 2016: 164-169.
- [16] ZHANG D, LIANG Z, YANG G, et al. A robust forgery detection algorithm for object removal by exemplar-based image inpainting[J]. *Multimedia Tools and Applications*, 2018, 77(10): 11823-11842.
- [17] ZHU X, QIAN Y, ZHAO X, et al. A deep learning approach to patch-based image inpainting forensics[J]. *Signal Processing: Image Communication*, 2018, 67: 90-99.
- [18] WANG X, WANG H, NIU S. An intelligent forensics approach for detecting patch-based image inpainting[J]. *Mathematical Problems in Engineering*, 2020, 2020(8): 1-10.
- [19] WANG X, NIU S, WANG H. Image inpainting detection based on multi-task deep learning network[J]. *IETE Technical Review*, 2021, 38(1): 149-157.
- [20] LU M, NIU S. A detection approach using LSTM-CNN for object removal caused by exemplar-based image inpainting[J]. *Electronics*, 2020, 9(5): 858.
- [21] BARNI M, COSTANZO A, NOWROOZI E, et al. CNN-based detection of generic contrast adjustment with JPEG post-processing[C]//2018 25th IEEE International Conference on Image Processing (ICIP). 2018: 3803-3807.
- [22] SHAN W, YI Y, HUANG R, et al. Robust contrast enhancement forensics based on convolutional neural networks[J]. *Signal Processing: Image Communication*, 2019, 71: 138-146.
- [23] ZHOU P, HAN X, MORARIU V I, et al. Learning rich features for image manipulation detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1053-1061.
- [24] BARNI M, BONDI L, BONETTINI N, et al. Aligned and non-aligned double JPEG detection using convolutional neural networks[J]. *Journal of Visual Communication and Image Representation*, 2017, 49: 153-163.
- [25] 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.
- [26] BUNK J, BAPPY J H, MOHAMMED T M, et al. Detection and localization of image forgeries using resampling features and deep learning[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017: 1881-1889.
- [27] 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.
- [28] 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. 2019: 9543-9552.
- [29] BARNI M, PHAN Q T, TONDI B. Copy move source-target disambiguation through multi-branch CNNs[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 1825-1840.
- [30] LI H, HUANG J. Localization of deep inpainting using high-pass fully convolutional network[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8301-8310.
- [31] WANG X, WANG H, NIU S. An image forensic method for AI inpainting using faster R-CNN[C]//International Conference on Artificial Intelligence and Security. 2019, 11634: 476-487.
- [32] LI A, KE Q, MA X, et al. Noise doesn't lie: towards universal detection of deep inpainting[C]//Proceedings of International Joint Conference on Artificial Intelligence. 2021: 786-792.
- [33] WU H, ZHOU J. IID-Net: image inpainting detection network via neural architecture search and attention[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(3): 1172 - 1185.
- [34] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
- [35] ZHANG D, CHEN X, LI F, et al. Seam-carved image tampering detection based on the cooccurrence of adjacent lbps[J]. Security and Communication Networks, 2020, 2020: 1-12.
- [36] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.
- [37] EVERINGHAM M, ESLAMI S, VAN GOOL L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.

[作者简介]



章登勇 (1980-), 男, 江西南昌人, 长沙理工大学副教授, 主要研究方向为多媒体信息安全、图像处理与模式识别。



文凰 (1996-), 女, 湖南望城人, 长沙理工大学硕士生, 主要研究方向为多媒体信息安全。



李峰 (1964-), 男, 湖南常德人, 长沙理工大学教授, 主要研究方向为多媒体信息安全、图像处理与模式识别。



曹鹏 (1982-), 男, 湖南常德人, 长沙理工大学讲师, 主要研究方向为图像处理与模式识别、图像取证。



向凌云 (1983-), 女, 湖南娄底人, 长沙理工大学副教授, 主要研究方向为信息安全、信息隐藏与数字水印、隐写分析、自然语言处理、模式识别与机器学习。



杨高波 (1973-), 湖南岳阳人, 湖南大学教授、博士生导师, 主要研究方向为图像/视频信息安全、多媒体通信、纹理压缩。



丁湘陵 (1981-), 男, 湖南株洲人, 博士, 湖南科技大学副教授, 主要研究方向为多媒体内容安全、图像/视频处理、密码分析。