

基于深度神经网络的图像修复取证算法

朱新山^{1,2**}, 钱永军¹, 孙彪^{1*}, 任超¹, 孙亚¹, 姚思如¹

¹天津大学电气自动化与信息工程学院, 天津 300072;

²中国科学院信息工程研究所信息安全国家重点实验室, 北京 100093

摘要 提出一种基于深度神经网络的图像修复取证算法,该算法可通过编码器网络自动提取图像修复遗留的痕迹特征,通过解码器网络预测像素类别,从而判断出图像是否经过修复篡改以及修复篡改的区域。同时,采用特征金字塔网络对解码器网络中的特征图进行信息补充。采用 MIT Place 数据集作为训练集,UCID 数据集作为测试集,对训练集和测试集分别使用了不同的修复篡改算法。实验结果表明,与其他图像修复取证算法相比,所提算法的修复区域定位更精准,处理速度更快,且对不同的修复篡改方法具有较好的稳健性和较强的泛化能力。

关键词 图像处理; 图像修复取证; 深度神经网络; 编码器网络; 解码器网络; 稳健性

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/AOS201838.1110005

Image Inpainting Forensics Algorithm Based on Deep Neural Network

Zhu Xinshan^{1,2**}, Qian Yongjun¹, Sun Biao^{1*}, Ren Chao¹, Sun Ya¹, Yao Siru¹

¹School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;

²State Key Laboratory of Information Security, Institute of Information Engineering,
Chinese Academy of Sciences, Beijing 100093, China

Abstract A novel image inpainting forensics algorithm based on the deep neural network is proposed, in which the vestigial features can be automatically extracted by the encoder network, the category of each pixel is predicted by the decoder network, and thus whether or not the image is with inpainting and falsification as well as the inpainted and falsified regions can be distinguished. Simultaneously, the feature pyramid network (FPN) is used to supplement the feature map in the decoder network. The MIT Place dataset is used as the training set and the UCID dataset as the testing set. In addition, the different inpainting and falsification algorithms are adopted for the training set and the testing set, respectively. The experimental results show that, compared with the other inpainting forensics algorithms of images, the proposed algorithm has a more accurate inpainting area and a faster processing speed. Moreover, it has relatively good robustness and strong generalization ability against different inpainting forensics algorithms.

Key words image processing; image inpainting forensics; deep neural network; encoder network; decoder network; robustness

OCIS codes 100.3020; 100.1455; 100.4996

1 引 言

随着互联网技术的快速发展以及数字图像获取设备的普及,数字图像信息的数量得到了快速增长;同时,由于图像处理和编辑软件的功能日益强大,非专业用户可方便地编辑或者篡改数字图像内容,因此图像信息的真实性和可靠性受到人们广泛关注。

为了解决这个问题,一些学者对数字图像篡改取证进行了研究。目前,针对特定图像篡改技术的取证算法主要包括双重 JPEG 压缩^[1]、中值滤波^[2]、复制粘贴^[3]等,但对图像修复取证的研究较少。

图像修复是一种新兴的图像处理算法,许多学者对其进行了深入研究。Criminisi 等^[4]提出一种基于图像块复制粘贴的修补方法,该方法采用基于

收稿日期: 2018-05-22; 修回日期: 2018-07-03; 录用日期: 2018-07-12

基金项目: 国家自然科学基金(61401303,51578189)、国家留学基金(201506255067)、信息安全国家重点实验室开放课题(2017-MS-11)、CCF 信息系统开放课题(CCFIS2018-02-04)、天津大学自主创新基金(2017XZY-0090,2018XZC-0033)

* E-mail: sunbiao@tju.edu.cn; ** E-mail: xszhu126@126.com

块的采样过程来实现对纹理和结构信息的填充; Jiao 等^[5]采用人工蜂群搜索技术来提高修复过程的计算效率,使计算时间减少一个数量级以上;苏畅等^[6]提出一种基于能量约束的自适应加权图像盲复原算法,该算法具有较高的峰值信噪比和结构相似度,能有效恢复高斯类模糊图像,增强图像细节分辨能力,提高图像的主观视觉效果;周程灏等^[7]提出一种空间变化图像复原方法,并分析了图像复原的基本模型;Elad 等^[8]提出一种基于形态分量分析和图像稀疏表示的修补方法,用于填补重叠原理和卡通图层中的孔;Xie 等^[9]提出一种基于深度学习的图像修补算法,其自动从图像中删除复杂元素且不需要修复区域的先验信息;Cai 等^[10]提出一种基于卷积神经网络(CNN)的盲修复方法,其可自动识别和删除损坏的图像区域并进行修复。图像修复的主要思想就是利用图像多个参考区域的信息对待修复区域进行填充或者合成,并尽可能使修复前后的图像在结构和纹理上保持一致。图像修复常被用于图像丢失信息的恢复、图像划痕去除和目标移除等^[11],但该操作不会留下明显的修复痕迹,这就给图像修复取证带来了严峻挑战。

国内外许多学者对图像修复取证算法进行了研究。2008年,Wu 等^[12]率先提出了一种针对样本合成修复的图像取证算法,该方法利用零连通特征筛选出可疑部分(即相似图像块对),并通过相似块间的模糊隶属度进行修复块的识别,但该方法需要人为给定选择区域,误警率较高且计算量很大。Bacchuwar 等^[13]提出一种基于跳跃式块匹配的改进算法,与文献^[12]算法相比,该算法的计算量明显减少,但仍需手动选择检测区域,且以上两种算法都对图像背景中的天空、草地等一致性区域产生严重误检。Chang 等^[14]提出一种基于多区域关联的自动取证算法,该方法在利用零连通特征搜索相似块对、向量滤波后,使用多区域关联识别修复块和基于权重变换的映射方法,加快了对可疑块的搜索,改进了文献^[12]和文献^[13]算法的不足,但基于权重变换的映射方法不能同时优化搜索范围和负载因子,因此限制了检测准确性的进一步提高。Liang 等^[15]提出一种基于最大零连通特征和碎像拼接检测的图像取证算法,其利用中心像素映射的快速搜索算法来搜索可疑区域,进一步优化检测算法,提高搜索速度,但基于中心映射的搜索算法通用性较差,不能适用于所有的图像修复篡改方式。

近年来,神经网络广泛应用于图像领域,如

图像分割、图像融合、图像分类等,但在图像篡改取证中应用较少。Chen 等^[16]提出一种基于深度神经网络的取证算法来对均值滤波篡改图像进行取证,与人为寻找特征的取证技术相比,该算法能显著提高检测性能。Yang 等^[17]提出一种拉普拉斯 CNN 算法来检测重获取图像,该算法在不同尺寸图像库的检测准确率超过 95%。Yu 等^[18]提出将 CNN 应用于多类反取证研究中,该算法的平均检测准确率可达 96.9%。由此可见,神经网络在图像修复取证方面有着广泛的应用前景。

图像修复取证技术就是根据特定特征找出图像中的异常相似块。传统的图像修复取证技术在提取特征的过程中存在以下几个未知问题:被识别图像篡改区域的大致范围、尺寸和数量;被识别图像所使用的篡改算法^[19];被识别图像是否经过后处理。而神经网络中的 CNN 是一种可以产生层级特征的视觉模型,其在自动提取图像特征方面取得了卓越成效。因此,本文提出一种基于深度 CNN 的图像修复取证方法。该方法采用包含卷积层、batch normalization(BN)层、修正线性单元(ReLU)层和池化层的编码器网络来实现图像特征的自动提取;采用包含上采样层、Merge 层和卷积层的解码器网络将编码器网络提取的特征图恢复至原图像大小,并采用特征金字塔网络^[20-21]对上采样过程中的特征图进行信息补充。该方法能对图像修复遗留的痕迹实现像素级检测,且能对任意尺寸的输入图像进行修复取证。实验结果表明,该方法对于目标图像是否经过修复篡改和篡改区域的定位均具有较好的预测性能,对不同的图像修复篡改方案具有稳健性,且计算速度更快。

2 图像修复取证

2.1 图像修复原理

图像修复是指对图像中的信息破损区域进行视觉上合理填充以恢复图像完整性的过程,并且观察者事后难以察觉曾经破损的区域。目前,大多数图像修复算法都是对 Criminisi 算法^[4]进行改进,如文献^[22]和文献^[23]的算法。Criminisi 算法的修复过程如图 1 所示,其中 Ω 为待修复区域, $\partial\Omega$ 为修复区域的边界, Φ 为图像中没有丢失信息的已知完整区域。图像修复的操作对象是图像块,Criminisi 算法的具体步骤如下:

1) 对图 1(a)中 $\partial\Omega$ 上的每一个点计算优先权,得到优先级最高的点 p ,然后选取以 p 为中心的图像块 Ψ_p 作为当前修复块,如图 1(b)所示;

2) 在图像已知区域 Φ 中搜索与 Ψ_p 最为匹配的图像块 Ψ_q , 如图 1(c) 所示;

3) 将 Ψ_p 中待填充的部分用最佳匹配块 Ψ_q 中

相应位置的像素代替;

4) 更新边缘, 如图 1(d) 所示, 重复步骤 1)~3), 直到待修复区域 Ω 修补完毕。

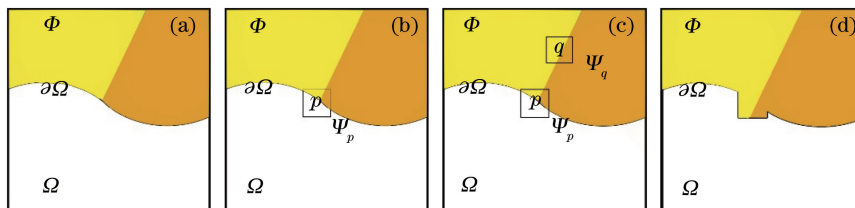


图 1 Criminisi 算法的图像修复过程。(a)确定修复块;(b)确定匹配块;(c)图像块修复;(d)更新边缘

Fig. 1 Image inpainting process by Criminisi algorithm. (a) Determining the inpainting block;

(b) determining the matching block; (c) inpainting the image block; (d) renewal the margin

2.2 图像修复取证

图像内容经过修复算法篡改后, 对图像进行的任何操作都会留下痕迹, 相应的图像修复取证算法可以通过分析目标图像的空间域特征或者频域特征性质的变化, 对图像进行修复取证。图 2(a) 为原图, 图 2(b) 为篡改后的图像, 图像修复取证就是检测出图 2(b) 中经过修复算法篡改的区域, 如图 2(c) 所示。图 2(c) 用不同的像素值表示图像内容是否经过修复篡改, 像素值 0 表示图像中未被修复篡改的区域, 像素值 1 表示图像中经过修复篡改的区域。

传统的图像修复取证算法对修复图像进行取证时, 一般先进行特征提取, 按照一定的区分规则对图

像内容进行判定, 初步选择出图像中的篡改区域, 然后结合一定的先验知识进行后处理操作, 更加精确地确定篡改区域, 最终输出图像修复取证的结果。传统算法的主要缺点是处理步骤多、时间长, 且检测准确率较低。而神经网络中的 CNN 是分层的网络结构, 浅层卷积部分可提取出图像的局部特征, 例如边缘和纹理特征等; 深层卷积部分可提取图像的全局特征, 将局部特征与全局特征相结合能完整地表达图像信息。因此, 基于 CNN 能自动提取图像特征, 构建了图像修复取证网络, 对修复图像进行自动判别, 输出如图 2(c) 所示的图像修复取证结果。



图 2 图像修复取证。(a)原图;(b)篡改后的图像;(c) CNN 修复取证后的图像

Fig. 2 Image inpainting forensics. (a) Original image; (b) falsified image; (c) image after CNN inpainting forensics

3 基于深度神经网络的图像修复取证

CNN 的感知域较大, 提取特征比较抽象, 这有助于图像分类, 但会丢失物体的具体轮廓和像素信息。同时, CNN 倾向于学习表达图像主要的内容特征, 忽略微弱的特征, 而图像修复篡改后遗留的痕迹较弱^[19], 因此在充分考虑图像修复原理和使用 CNN 进行图像修复取证存在问题的基础上, 构建了一个适合图像修复取证的全卷积网络结构。全卷积网络是训练出一个端到端、点到点的网络, 从抽象的特征中恢复出每个像素的类别。为了使网络能够识别篡改痕迹而不受内容影响, 通过全卷积网络的编

码器网络和解码器网络将输出结果恢复至输入图像大小, 同时给输出图像中的每个像素点一个类别标签(1 或者 0, 1 表示该像素点在篡改区域, 0 表示该像素点不在篡改区域), 以实现像素类别检测。此外, CNN 提取图像特征的过程会造成图像内容损失, 故采用了特征金字塔网络^[20-21]对全卷积网络中的特征图进行信息补充, 填充图像的语义信息。

3.1 网络结构设计

为了实现取证区域的预测, 采用包含编码器网络和解码器网络的全卷积网络结构进行篡改痕迹识别, 如图 3 所示。其中: 编码器网络部分用于自动提取图像特征, 输出尺寸为输入图像尺寸 1/5 的高

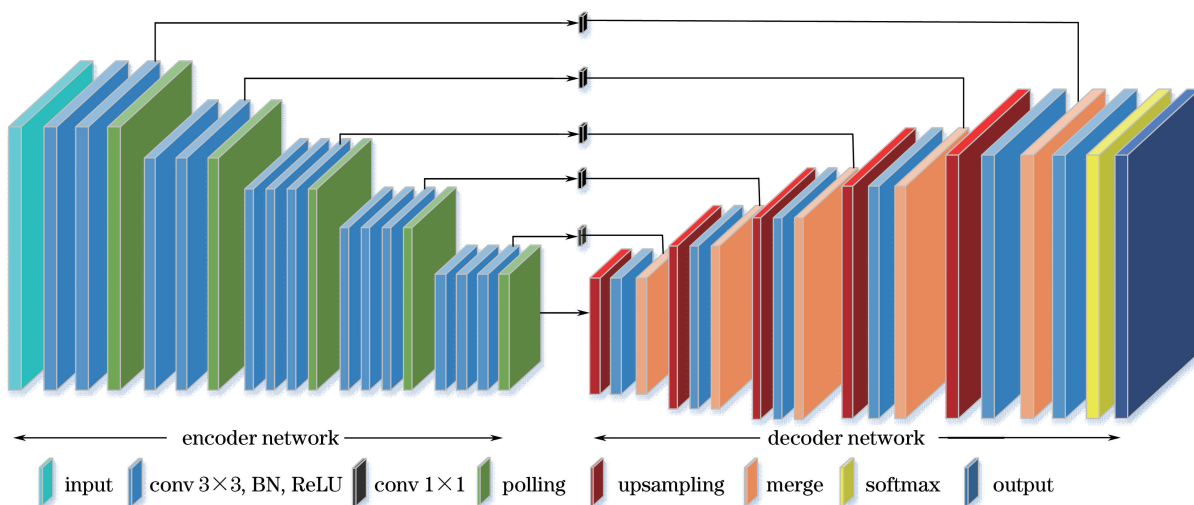


图 3 网络结构示意图

Fig. 3 Structural diagram of network

维特征图;解码器网络部分用于将提取的高维特征图恢复至输入图像大小。然后把解码器网络的输出送到多类 softmax 分类器中,多类 softmax 分类器可为每个像素点产生类别概率,即每个像素点是否属于篡改区域的概率。

编码器网络中的每个编码器都先通过卷积层生成一组特征图。卷积层通过局部感受野和权值共享实现卷积,提取目标图像特征。局部感受野是指后一层单个神经元通过卷积核只连接前一层输入的某个区域;权值共享是指每个神经元都采用相同的卷积核参数,以降低网络学习的参数量。卷积层可通过选择不同的卷积核来提取待取证图像的多个特征。卷积层后面紧跟 BN 层,由于在网络训练的过程中,每层的输入都在不断变化,导致网络训练难度增大,但 BN 层可以解决这个问题,同时有助于抑制网络过拟合。然后再接 ReLU 层,ReLU 激活函数使得网络具有稀疏性,并减少参数的相互依存关系,从而有助于对抗网络训练中的过拟合。之后紧跟尺寸为 $2 \text{ pixel} \times 2 \text{ pixel}$,步长为 2 的最大池化层,通过保留最大值可减少 75% 的冗余,降低运算复杂度,其输出下采样特征图尺寸为输入特征图的 $1/2$ 。

解码器网络中的每个解码器都先通过上采样层,对输入特征图进行特征映射。上采样层采用了基于双线性的上采样插值,其核心思想是在两个方向上分别进行一次线性插值,插值算法放大后的图像质量较高,不会出现像素值不连续的情况,因此可

以产生粗糙的特征图。上采样层后面接尺寸为 $3 \text{ pixel} \times 3 \text{ pixel}$,步长为 1 的卷积。由于 CNN 的卷积和下采样操作会导致图像信息损失,直接将卷积得到的高阶特征图进行上采样后的输出图像比较模糊,与前面对应编码器的低层特征图相融合构成特征金字塔网络能够有效地补充卷积和池化过程中的图像信息损失,因此本文算法采用融合层。但编码器和解码器输出特征图的通道数不同,所以不能直接通过融合层进行融合,应先使编码器的输出特征图 X_i 经过 $\mu(X_i)$ 操作,即尺寸为 $1 \text{ pixel} \times 1 \text{ pixel}$ 的卷积,改变编码器输出特征图的通道数,如图 3 的中上部分所示,然后使输出 $\mu(X_i)$ 与对应解码器的输出特征图 Y_{i-1} 经过 $\tau[\cdot]$ 按通道进行融合得到 Y_i ,细化上采样得到的特征图,补充编码器网络的卷积和下采样操作带来的图像信息损失:

$$Y_i = \tau [Y_{i-1}, \mu(X_i)], \quad (1)$$

式中: X_i 为第 i 个编码器的输出特征图; Y_i 为第 i 个解码器的输出特征图。最终编码器网络的输出特征图经过每个解码器的上采样层、卷积层和融合层后,输出与原图一样大小的特征图,该特征图保留了原始输入图像的空间信息。随后将其送入 softmax 层,softmax 层可以对输入特征图的每一个像素点产生一个分类概率,计算每个像素的 softmax 分类损失,其相当于每一个像素点对应一个训练样本,故本文算法可检测出图像修复遗留的操作痕迹,从而准确定位图像修复篡改的区域。

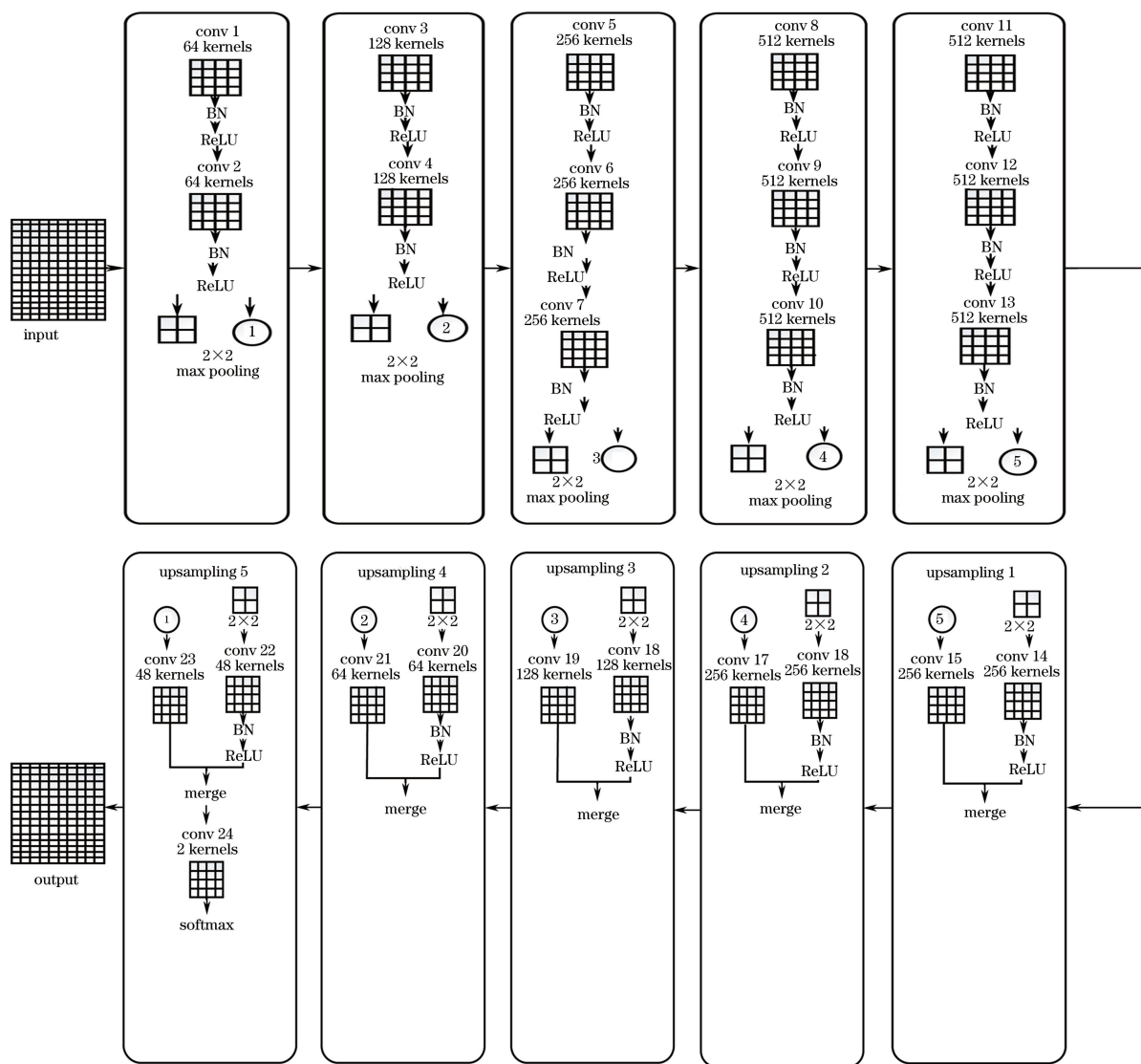


图 4 网络结构原理图

Fig. 4 Structural diagram of network

3.2 网络结构参数说明

由图 3 可知,全卷积网络包括编码器网络和解码器网络,具体网络结构参数如图 4 所示。

编码器网络包含 5 个编码器,其由 visual geometry group(VGG)的前 13 层组成,用于自动提取目标图像的修复篡改特征。前 2 个编码器分别具有 2 个卷积层:BN 层和 ReLU 层;后 3 个编码器分别具有 3 个卷积层:BN 层、ReLU 层,以及每个编码

器后面跟一个最大池化层,编码器的特征图数量和特征图尺寸如表 1 所示。

解码器网络包括 5 个解码器,其由卷积层、上采样层和融合层组成,用于将提取的特征恢复至输入图像大小,实现像素类别检测。前 4 个解码器分别具有 2 个卷积层、1 个上采样层和 1 个融合层,第 5 个解码器有 3 个卷积层、1 个上采样层和 1 个融合层。解码器的特征图数量和特征图尺寸如表 1 所示。

表 1 全卷积网络结构参数

Table 1 Structural parameters of convolution-wide network

Type of network	Encoder network					Decoder network				
Number of feature maps	64	128	256	512	512	256	256	128	64	48
Feature size	256×256	128×128	64×64	32×32	16×16	16×16	32×32	64×64	128×128	256×256

4 实验结果与分析

4.1 实验数据与评价指标

本文算法采用基于 Tensorflow 后端的 Keras 框架,网络采用随机梯度下降法,动量为 0.95,批量大小为 16。同时,为了使网络能快速地沿着梯度下降方

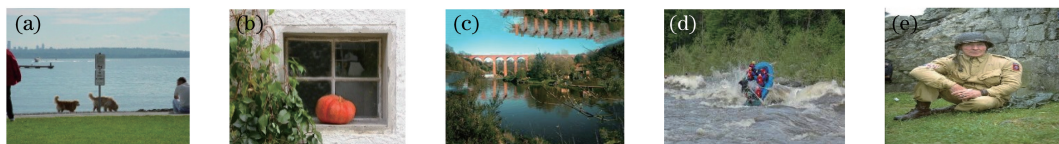


图 5 测试集的部分经典图像

Fig. 5 Partial classical images in testing set

1) 训练集数据和验证集数据。先从 MIT Place 数据集中随机抽取 1 万张彩色图像,然后通过旋转和镜像操作得到 5 万张图像,其中 80% 作为训练集数据,20% 作为验证集数据,最后使用 Criminisi 修复算法对数据集图像进行篡改(图 6 中绿色区域为篡改区域)。篡改区域分为两种:如图 6(a)~(c)所示的规则篡改区域和图 6(d)~(f)所示的不规则篡改区域。规则篡改区域包括矩形篡改区域和圆形篡改区域,并按区域面积分为三种,即篡改区域面积分别占图像总面积的 5%、10%、20%;不规则篡改区域包括单篡改区域和多篡改区域,面积随机选取,其范围为 0~50%。

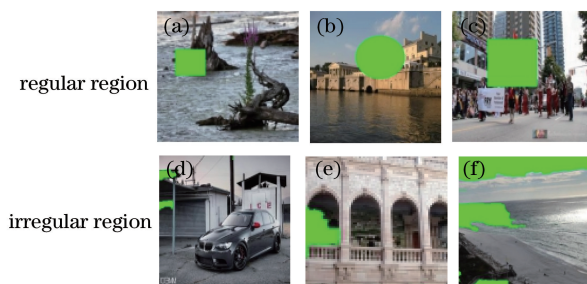


图 6 不同篡改区域面积占比的篡改图像。(a) 5%;(b) 10%;(c) 20%;(d) 0~5%;(e) 10%~30%;(f) 30%~50%

Fig. 6 Falsified images with different area ratios of falsification regions. (a) 5%; (b) 10%; (c) 20%; (d) 0-5%; (e) 10%-30%; (f) 30%-50%

2) 测试集数据。在图 5 所示的经典图像和 UCID 数据集中随机抽取 2000 张彩色图像作为测试集数据。使用图 5 所示的经典图像可以实现本文算法与传统方法的取证性能对比,使用 UCID 数据集可以从整体上评价本文算法的取证性能和稳健性。首先移除图 5 所示的测试图像的语义对象,然后使用 Criminisi 算法进行修复。UCID 测试集图

向更新参数,加快算法收敛速度,得到性能较好的参数,需要在一定时间内不断减少网络的学习率,因此本文网络的初始学习率设为 0.004,权重衰减设为 10^{-4} 。训练集数据和验证集数据使用 MIT Place 数据集,测试集数据使用 UCID 数据集^[24]和文献[14]中的图像,部分测试集图像如图 5 所示。

像分别使用 Criminisi 算法和 Shift-map 算法^[25]修复,其修复区域的选取和训练集数据一样,其中 5%、10%、20% 的规则篡改图像各 200 张,共 1200 张,不规则篡改图像共 400 张,未篡改图像 400 张。

3) 标签数据。标签图像是与测试图像尺寸相同的二值图,像素值为 1 的区域对应于测试图像的篡改区域,像素值为 0 的区域对应于测试图像的非篡改区域。

4) 评价指标。为了评估本文算法的取证性能,采用以下 3 种指标进行检测性能评价:真正率 (TPR)、假正率 (FPR),以及确定篡改区域需要花费的时间 T 。TPR 和 FPR 可表示为

$$R_{TP} = \frac{|p \cap q|}{|p|}, \quad (2)$$

$$R_{FP} = \frac{|q \cap h|}{|h|}, \quad (3)$$

式中: $|\cdot|$ 为集合的势; p 为原有的修复区域; h 为原有的非修复区域; q 为检测出的修复区域。

4.2 不同方案的检测性能比较

为了评估本文算法的检测性能,将实验结果与文献[14]和文献[15]的图像修复取证结果进行比较。实验数据采用图 5 所示的经典图像以及 UCID 数据集中的 200 张图像,实验结果如图 7、表 2 和表 3 所示。由表 2 可得 3 种算法分别在图 5(a)~(e)的 TPR 和 FPR,其中文献[14]算法在不同图像上的 TPR 和 FPR 差值均较大,如在图 5(a)和图 5(e)上的 TPR 值分别为 66.55% 和 0,在图 5(c)上的 FPR 为 30.09%;文献[15]算法在不同图像上的 TPR 值较接近,所有图像的 TPR 值均为 87% 左右,但 FPR 值相差较大,如图 5(b)的 FPR 值为 0,图 5(e)的 FPR 值最大,为 14.42%。上述实验结果表明文献[14]和文献[15]算法的取证性能依

表 2 不同算法对经典图像的 TPR 和 FPR 比较(单位: %)

Table 2 TPR and FPR of classical images by different algorithms (unit: %)

Image No.	Method in Ref. [14]		Method in Ref. [15]		Proposed method	
	TPR	FPR	TPR	FPR	TPR	FPR
Fig. 5(a)	66.55	1.24	88.53	0.2	95.03	4.87
Fig. 5(b)	90.42	4.13	85.01	0	98.03	0.16
Fig. 5(c)	86.04	30.09	83.15	0.86	96.78	3.22
Fig. 5(d)	92.42	6.75	92.16	0.27	98.99	0.09
Fig. 5(e)	0	10.09	89.29	14.42	97.69	0.15

表 3 不同算法对 UCID 数据集图像的 TPR 和 FPR 比较(单位: %)

Table 3 TPR and FPR of images in UCID dataset by different algorithms (unit: %)

Image Num.	Method in Ref. [14]		Method in Ref. [15]		Proposed method	
	TPR	FPR	TPR	FPR	TPR	FPR
200	87.35	14.7	94.75	10.8	95.6	0.8

赖于图像的纹理和结构类型。而本文算法在所有图像中均取得了较理想的效果,具有较高的 TPR 值与较低的 FPR 值。表 3 为 3 种算法分别在 200 张 Criminisi 算法修复图像上的取证结果,由表 3 数据可知,本文算法的 TPR 值与 FPR 值均优于文献 [14] 和文献 [15] 算法。图 7 所示为 3 种算法在图 5(b)、(e) 的检测结果。从检测结果来看,与文献 [14] 和文献 [15] 算法相比,本文算法可以较准确地确定修复区域的位置和形状,且检验的误警图像块较少。综上所述,本文算法的取证性能整体优于文献 [14] 和文献 [15] 算法的取证性能。

4.3 本文算法的整体性能评价

为了全面评估本文算法对不同尺寸修复图像的取证稳健性,在 UCID 数据集进行了取证实验。UCID 数据集的图片具有不同的尺寸,如 $384 \text{ pixel} \times 512 \text{ pixel}$ 等。为了验证本文算法对不同篡改区域检测性能的稳健性,在 UCID 数据集图片上随机产生了任意数量的篡改区域; MIT Place 训

练集的修复篡改区域是由 Criminisi 算法形成,而为了验证本文算法对不同修复算法的稳健性,分别对 UCID 测试集的篡改区域使用了 Criminisi 算法和 Shift-map 算法进行修复篡改。表 4 所示为本文算法对采用 Criminisi 算法修复篡改图像的检测结果,表 5 所示为本文算法对采用 Shift-map 算法修复篡改图像的检测结果。

由表 4 可知,本文提出的图像修复取证算法在规则修复篡改区域和不规则修复篡改区域的取证中均具有较好的检测性能,评价指标 TPR 的最大值为 98.64%, 最小值为 90.43%, FPR 的最小值为 0.15%, 最大值为 1.74%, 且平均检测时间为 1.9 s。由于测试集中待取证图像的篡改区域分别为规则篡改区域和不规则篡改区域,网络不能通过学习篡改区域的形状特征进行修复取证,也说明网络学习的特征不具有普遍性,这也说明本文算法能够有效提取修复图像遗留的操作痕迹来对目标图像进行取证,实现像素类别检测。

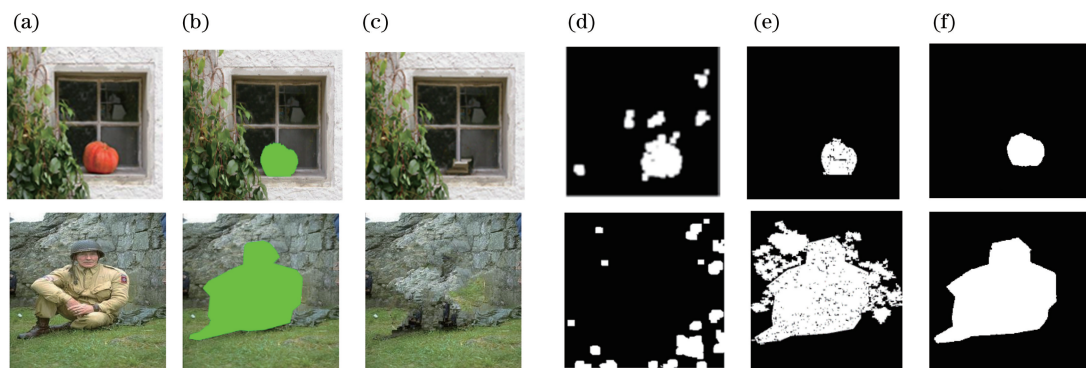


图 7 修复区域检测结果。(a)原始图像;(b)掩模图像;(c)修复图像;(d)文献[14]算法的检测结果;
(e)文献[15]算法的检测结果;(f)本文算法的检测结果

Fig. 7 Detection results of inpainting regions. (a) Original image; (b) mask image; (c) inpainted image; (d) detection results by algorithm in Ref. [14]; (e) detection results by algorithm in Ref. [15]; (f) detection results by proposed method
1110005-7

表 4 采用 Criminisi 算法修复图像的检测结果

Table 4 Detection results of images inpainted by Criminisi algorithm

Parameter	Mask size /%	TPR /%	FPR /%	T /s
Regular region	5	98.64	0.34	1.8
	10	97.84	0.24	1.9
	20	96.90	0.15	1.8
Irregular region	0-10	90.43	0.4	2.1
	10-30	94.88	1.08	2.0
	30-50	96.96	1.74	1.8
Mean	—	95.94	0.674	1.9

表 5 采用 Shift-map 算法修复图像的检测结果

Table 5 Detection results of images inpainted by Shift-map algorithm

Parameter	Mask size /%	TPR /%	FPR /%	T /s
Regular region	5	93.29	0.29	1.9
	10	92.48	0.19	1.8
	20	89.95	0.25	2
Irregular region	0-10	80.09	0.32	2.1
	10-30	80.36	2.18	2.2
	30-50	73.79	6.17	2.0
Mean	—	84.99	1.57	2

由表 5 可知,当测试集图像和训练集图像分别使用不同修复算法进行篡改时,本文算法仍然具有较好的检测性能。MIT Place 训练集数据仅采用了 Criminisi 算法进行图像修复篡改,当采用 Shift-map 算法修复图像时,TPR 值有所下降,但仍然具有较高的准确率,这也说明本文算法对于不同的图像修复篡改算法具有较好的稳健性和较强的泛化能力。

5 结 论

提出一种基于深度神经网络的图像修复取证算法,该算法集成了编码器网络和解码器网络,其中编码器网络由卷积层、BN 层、ReLU 层和池化层组成,可自动提取图像修复遗留的特征;解码器网络包括上采样层、卷积层和融合层,可实现像素类别的预测,从而判断图像是否经过修复篡改以及图像修复篡改的区域。将卷积层代替全连接层,实现对任意尺寸图像的修复取证,采用特征金字塔网络对特征图进行信息补充,填充图像的语义信息。本文算法已在大量图片上进行了测试,并与最先进的修复取证方法进行了比较。实验结果表明,基于深度神经网络的取证方案可以自动学习如何检测图像的修复篡改操作,而无需考虑特征提取和分类设计。该算法对不同尺寸、数量和形状的篡改图像均具有较好

的取证性能,且检测时间少,具有较好的稳健性和较强的泛化能力。

参 考 文 献

- [1] Taimori A, Razzazi F, Behrad A, *et al.* A novel forensic image analysis tool for discovering double JPEG compression clues [J]. *Multimedia Tools and Applications*, 2017, 76(6): 7749-7783.
- [2] Rhee K H. Median filtering detection using variation of neighboring line pairs for image forensics [J]. *Journal of Electronic Imaging*, 2016, 25(5): 053039.
- [3] Li J W, Yang F, Lu W, *et al.* Keypoint-based copy-move detection scheme by adopting MSCRs and improved feature matching [J]. *Multimedia Tools and Applications*, 2017, 76(20): 20483-20497.
- [4] Criminisi A, Perez 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] Jiao A S M, Tsang P W M, Poon T C. Restoration of digital off-axis Fresnel hologram by exemplar and search based image inpainting with enhanced computing speed [J]. *Computer Physics Communications*, 2015, 193: 30-37.
- [6] Su C, Fu T J, Zhang X X, *et al.* Adaptively-weighted blind image restoration algorithm based on energy constraint [J]. *Acta Optica Sinica*, 2018, 38(2): 0210001.
苏畅, 付天骄, 张星祥, 等. 基于能量约束的自适应加权图像盲复原算法 [J]. *光学学报*, 2018, 38(2): 0210001
- [7] Zhou C H, Wang Z L, Liu S K. Method of image restoration directly based on spatial varied point spread function [J]. *Acta Optica Sinica*, 2017, 37(1): 0110001.
周程灏, 王治乐, 刘尚阔. 基于空间变化点扩展函数的图像直接复原方法 [J]. *光学学报*, 2017, 37(1): 0110001.
- [8] Elad M, Starck J L, Querre P, *et al.* Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA) [J]. *Applied and Computational Harmonic Analysis*, 2005, 19(3): 340-358.
- [9] Xie J, Xu L, Chen E. Image denoising and inpainting with deep neural networks [C] // *International Conference on Neural Information Processing Systems*, 2012: 341-349.
- [10] Cai N, Su Z H, Lin Z N, *et al.* Blind inpainting using the fully convolutional neural network [J]. *The Visual Computer*, 2017, 33(2): 249-261.

- [11] Ružić T, Pižurica A. Context-aware patch-based image inpainting using markov random field modeling [J]. Proceedings of the IEEE, 2015, 24(1): 444-456.
- [12] Wu Q, Sun S J, Zhu W, *et al.* Detection of digital doctoring in exemplar-based inpainted images [C]. International Conference on Machine Learning and Cybernetics, 2008, 3: 1222-1226.
- [13] Bacchuwar K S, Aakashdeep, Ramakrishnan K R. A jump patch-block match algorithm for multiple forgery detection[C]. International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing, 2013: 723-728.
- [14] 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.
- [15] Liang Z S, Yang G B, Ding X L, *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.
- [16] Chen J S, Kang X G, Liu Y, *et al.* Median filtering forensics based on convolutional neural networks[J]. Proceedings of the IEEE, 2015, 22(11): 1849-1853.
- [17] Yang P, Ni R, Zhao Y. Recapture image forensics based on Laplacian convolutional neural networks [C]. International Workshop on Digital Watermarking, 2016, 10082: 119-128.
- [18] Yu J, Zhan Y, Yang J, *et al.* A multi-purpose image counter-anti-forensic method using convolutional neural networks [C]. 15th International Workshop, 2016: 3-15.
- [19] Zhu X S, Qian Y J, Zhao X F, *et al.* A deep learning approach to patch-based image inpainting forensics [J], Signal Processing Image Communication. 2018, 67: 90-99.
- [20] Lin T Y, Dollár P, Girshick R, *et al.* Feature Pyramid Networks for Object Detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 936-944.
- [21] He K, Gkioxari G, Dollár P, *et al.* Mask R-CNN [C]. IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [22] Wang H, Cai Y, Liang R, *et al.* Exemplar-based image inpainting using structure consistent patch matching [C]. International Conference on Intelligence Science and Big Data Engineering, 2015, 269: 401-410.
- [23] Chen Z H, Dai C, Jiang L, *et al.* Structure-aware image inpainting using patch scale optimization [J]. Journal of Visual Communication and Image Representation, 2016, 40: 312-323.
- [24] Schaefer G, Stich, M. UCID An uncompressed colour image database [C]. Conference on Storage and Retrieval Methods and Applications for Multimedia, 2003, 5307: 472-480.
- [25] Pritch Y, Kav-Venaki E, Peleg S. Shift map image editing [C]. IEEE International Conference on Computer Vision, 2009, 30: 151-158.