基于级联卷积神经网络的图像篡改检测算法

毕秀丽[①] 魏 杨[①] 肖 斌*[①] 李伟生[①] 马建峰[②]

[①](重庆邮电大学计算智能重点实验室 重庆 400065)

[②](西安电子科技大学网络与信息安全学院 西安 710071)

摘 要：基于卷积神经网络的图像篡改检测算法利用卷积神经网络的学习能力可以实现不依赖于单一图像属性的图像篡改检测，弥补传统图像篡改检测方法依赖单一图像属性、适用度不高的缺陷。利用深层多神经元的单一网络结构的图像篡改检测算法虽然可以学习更高级的语义信息，但检测定位篡改区域效果并不理想。该文提出一种基于级联卷积神经网络的图像篡改检测算法，在卷积神经网络所展示出来的普遍特性的基础上进一步探究其深层次的特性，利用浅层稀神经元的级联网络结构弥补以往深层多神经元的单一网络结构在图像篡改检测中的缺陷。该文提出的检测算法由级联卷积神经网络和自适应筛选后处理两部分组成，级联卷积神经网络实现分级式的篡改区域定位，自适应筛选后处理对级联卷积神经网络的检测结果进行优化。通过实验对比，该文算法展示了较好的检测效果，且具有较高的鲁棒性。

关键词：图像篡改检测；级联卷积神经网络；浅层稀神经元；级联网络结构；自适应筛选后处理

中图分类号：TP309.2 **文献标识码**：A **文章编号**：1009-5896(2019)12-2987-08

DOI: 10.11999/JEIT190043

Image Forgery Detection Algorithm Based on Cascaded Convolutional Neural Network

BI Xiuli[①] WEI Yang[①] XIAO Bin[①] LI Weisheng[①] MA Jianfeng[②]

[①](*Chongqing Key Laboratory of Computational Intelligence*, *Chongqing University of Posts and Telecommunications*, *Chongqing* 400065, *China*)

[②](*School of Computer Science and Technology*, *Xidian University*, *Xi'an* 710071, *China*)

Abstract: The image forgery detection algorithm based on convolutional neural network can implement the image forgery detection that does not depend on a single image attribute by using the learning ability of convolutional neural network, and make up for the defect that the previous image forgery detection algorithm relies on a single image attribute and has low applicability. Although the image forgery detection algorithm using a single network structure of deep layers and multiple neurons can learn more advanced semantic information, the result of detecting and locating forgery regions is not ideal. In this paper, an image forgery detection algorithm based on cascaded convolutional neural network is proposed. Based on the general characteristics exhibited by convolutional neural network, and then the deeper characteristics are further explored. The cascaded network structure of shallow layers and thin neurons figures out the defect of the single network structure of deep layers and multiple neurons in image forgery detection. The proposed detection algorithm in this paper consists of two parts: the cascade convolutional neural network and the adaptive filtering post-processing. The cascaded convolutional neural network realizes hierarchical forgery regions localization, and then the adaptive filtering post-processing further optimizes the detection result of the cascaded convolutional neural network. Through experimental comparison, the proposed detection algorithm shows better detection results and has higher robustness.

收稿日期：2019-01-15；改回日期：2019-04-19；网络出版：2019-05-21

*通信作者： 肖斌 xiaobin@cqupt.edu.cn

基金项目：国家自然科学基金(61572092, U1401252)，国家重点研发计划基金(2016YFC1000307-3)

Foundation Items: The National Natural Science Foundation of China (61572092, U1401252), The National Science & Technology Major Project (2016YFC1000307-3)

Key words: Image forgery detection; Cascaded convolutional neural network; Shallow layers and thin neurons; Cascaded network structure; Adaptive filtering post-processing

1 引言

随着计算机技术和数字媒体技术的快速发展，大量的图像编辑软件的出现使得人们可以轻易地篡改图像的内容信息，这使得图片真实性急剧下降，导致图像的可信度严重降低。这些频繁出现的篡改图像已经在军事、司法、传媒等领域引起了严重的不良后果，同时也引发了研究学者们对图像篡改问题的关注。在研究学者们已有的研究范围里，图像内容篡改操作总体被分为两类：(1)复制粘贴(copy-move)篡改；(2)剪切组合(splicing)篡改。复制粘贴篡改是指将一幅图像里的某一部分内容复制粘贴到同一幅图的另外一个部分，以达到掩盖或者增加图像内容的目的。剪切组合篡改是指将一幅图像的一个或几个区域拷贝到另一幅图像中以生成一幅新的图像。本文主要针对剪切组合篡改进行研究，所提出的方法实现剪切组合篡改检测。在剪切组合篡改中，由于被篡改图像和篡改区域来自不同的图像，而不同图像之间一定存在属性差异，利用这些属性差异是实现篡改的检测和篡改区域的定位的主要依据。传统的剪切组合篡改检测方法根据其依赖的图像属性的类型大体可被分为4类：(1)基于图像本质属性的检测方法；(2)基于成像设备属性的检测方法；(3)基于图像压缩属性的检测方法；(4)基于图像哈希的检测方法。以上检测算法普遍利用一种特定的图像属性，故这些算法的适用度不高，在一些情况下不能成功完成检测。例如：(1)当图像在被剪切组合篡改后再执行一些隐藏的后处理(例如：直方图均衡化，整体模糊)后，基于图像本质属性的检测方法会失效；(2)当被篡改图像和篡改图像本身的噪声信息强度不足时，基于成像设备属性的检测方法会失效；(3)基于JPEG图像压缩属性的检测方法只能检测JPEG格式的图像；(4)基于图像哈希技术的检测方法需要依赖原始图像的哈希值，当哈希值未知或被破坏时，则检测方法会失效。文献[1]是一种基于成像设备属性的检测算法，其利用小波滤波来提取局部图像噪声方差模型以实现图像篡改区域的定位。文献[2–4]皆是基于图像压缩属性的检测算法，其中文献[2]通过将可疑图像重压缩为JPEG图像格式，并从原始图像中减去重压缩后的图像以实现篡改检测；文献[3]利用图像的DCT系数直方图来完成图像中的不一致检测从而实现篡改区域的定位；而文献[4]则从图像的DCT系数中进行非对齐双量化检测以实现最终的篡改定位。

近年来，基于卷积神经网络的图像篡改检测算法利用卷积神经网络的学习能力可以实现不依赖于单一图像属性的图像篡改检测，弥补传统图像篡改检测方法依赖单一图像属性、适用度不高的缺陷。在文献[5]中，卷积神经网络被首次应用于图像篡改检测，此算法可用于判断当前图像是否被篡改，但不能定位篡改区域的具体位置。文献[6]提出了一种基于图像块的检测算法，但仅能大致定位图像中的篡改区域。为实现篡改区域的像素级定位，在文献[7,8]中都使用了不重叠的图像块作为神经网络的输入进行判断。文献[9]利用图像的拍摄信息元数据作为监督信号来判断图像内容信息是否一致。以上利用深层多神经元的单一网络结构的图像篡改检测算法虽然可以学习更高级的语义信息，但检测定位篡改区域效果并不理想。

本文基于卷积神经网络的特性及设计思想，进一步研究并提出浅层稀神经元的级联网络结构以代替深层多神经元的单一网络结构，通过对困难样本特征进行再学习的方式，弥补以往利用深层多神经元的单一网络结构的图像篡改检测算法的缺陷。本文提出的检测算法由级联卷积神经网络和自适应筛选后处理两部分组成，级联卷积神经网络学习图像中篡改与非篡改区域之间的多种属性差异，实现分级式的篡改区域定位，自适应筛选后处理对级联卷积神经网络的检测结果进行精细优化。

本文的结构安排如下：第2节介绍级联卷积神经网络及自适应筛选后处理的具体步骤；第3节描述实验细节及与其它几种算法的对比实验效果；第4节总结全文。

2 基于级联卷积神经网络的图像篡改检测算法

基于级联卷积神经网络的图像篡改检测算法由两部分组成：级联卷积神经网络和自适应筛选后处理。本算法的检测流程如图1所示，级联卷积神经网络由3级浅层网络组成，首先，第1级粗筛网络会对图像中篡改区域进行粗筛定位，然后，第2级粒提网络进一步地对粗筛定位的篡改区域进行粒度级定位，最后，第3级精辨网络在粒度级定位的篡改区域上精细定位得到疑似篡改区域。自适应筛选后处理对级联网络输出的疑似篡改区域进行优化以输出最终的检测结果。基于级联卷积神经网络的图像篡改检测示例如图2所示，其中图2(a)为待检图像，图2(b)为通过粗筛网络检测得到的粗筛定位的篡改区域(f-out)，图2(c)为通过粒提网络检测得到的粒度级定位的篡改区域(s-out)，图2(d)为通过精

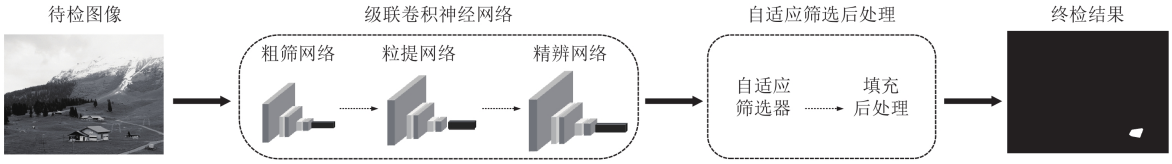
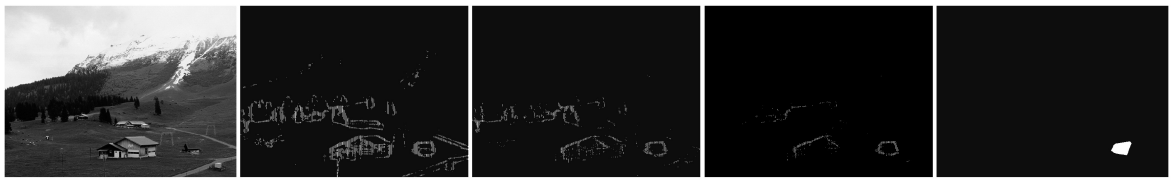


图 1 基于级联卷积神经网络算法的检测流程



(a) 待检图像　(b) 粗筛结果　(c) 粒提结果　(d) 精辨结果　(e) 终检结果

图 2 基于级联卷积神经网络算法的检测示例

辨网络检测得到的疑似篡改区域(t-out)，图2(e)为通过自适应筛选后处理筛选得到的最终的检测结果(AF-out)。

2.1 级联卷积神经网络

当前的卷积神经网络结构设计主要存在以下两大趋势：(1)网络结构更深；(2)节点神经元更多。以VGG[10]和ResNet[11]为例，前者作为2014年ImageNet大规模识别挑战赛分类项目冠军，其具有13层卷积层加上3层全连接层的网络结构，后者在2015年被提出后，获得多项比赛冠军，其具有超过50层卷积层的网络深度，两者都为深层网络且节点神经元较多，同时也都具备突出的特征提取能力。但深层及多神经元的网络结构，也代表此网络的前向传播及误差的反向传播更慢，进而网络的训练及测试所耗费的时间也更长。因此基于以上问题，本文提出了级联卷积神经网络，将深层多神经元的单一网络结构转化为浅层稀神经元的级联网络结构。同时由于级联网络结构的灵活性，通过向不同层级的网络传递不同的训练数据可以使其学习到特定的特征，并进一步利用这些特征来对图像篡改信息进行判断。级联卷积神经网络由粗筛网络、粒提网络以及精辨网络3级子网络串联组成，其检测流程如下：(1)首先将待检图像划分为一系列重叠的图像块，然后利用粗筛网络对每个图像块按序进行判别检测，得到的检测结果如图2(b)所示，该检测结果存在较多误检；(2)进一步地，将粗筛结果(图2(b))中可疑篡改区域以像素点为中心分别取出所有图像块，并利用粒提网络依次对每个图像块进行检测分类，其检测结果(图2(c))相较于粗筛结果可过滤掉较多的误检区域；(3)为最优化级联卷积神经网络的检测效果，利用精辨网络对粒提结果作进一步地精确定位，具体操作与步骤(2)相同，其检测结果如图2(d)所示。

2.1.1 粗筛网络

在级联卷积神经网络中，首先使用粗筛网络粗略地提取篡改区域与非篡改区域之间的特征差异。粗筛网络结构如图3所示，其包含4层卷积层及2层全连接层，图中数字表示下方对应层的神经元个数，例如：32即代表第1层卷积层拥有32个神经元。卷积操作如式(1)所示

$$y_i = w_i \times x + b_i \tag{1}$$

其中，x为输入，w_i为第i个神经元(也称为卷积核)的权重值，b_i为第i个神经元的偏置项，y_i为第i个神经元的输出。

在粗筛网络中，前两层卷积层中卷积步长都为1，在每层卷积后都使用了最大池化操作来对特征信息进行降维，并在降维后使用批归一化操作[12]对

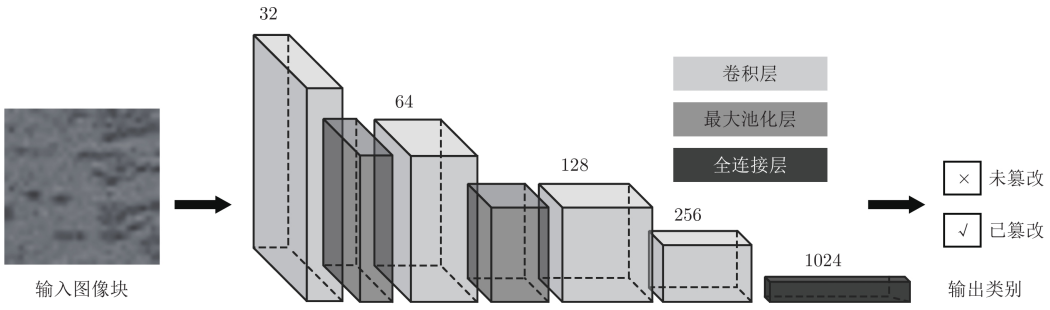


图 3 粗筛网络结构

数据进行归一化处理以便网络训练时更易收敛；在后两层卷积层中都使用了步长为2的卷积以替代最大池化层，其目的是为了保持特征的空间信息，并在卷积后同样使用批归一化操作对数据进行归一化处理，最后利用激活函数ReLU[13]对上述结果进行非线性拟合。这样粗筛网络既可以降维特征信息的同时又可以保留其空间位置信息，故可更有效地对图像中的篡改区域进行定位。在粗筛网络进行训练之前，需产生对应的训练数据，首先沿着篡改图像中篡改区域的边缘取大小为$P_{\mathrm{f}}\times P_{\mathrm{f}}$的块，并将这些块的标签设为已篡改，之后在对应的原始图像上的相同位置取同样大小的块，同时将这些块的标签设为非篡改。然后将所得训练数据输入到粗筛网络中进行训练并最终得到检测模型。粗筛网络在经过训练后具备了鉴别篡改与非篡改区域之间的属性差异的能力，在输入1张图像块后，其便会输出预测的图像块的类别：已篡改或非篡改。粗筛网络的检测结果如图2(b)所示，尽管在粗筛定位的篡改区域(f-out)中包含篡改区域，但却出现较多误检。

2.1.2 粒提网络

粒提网络结构如图4所示，前3层卷积中卷积步长都为1，在每层卷积后都使用了最大池化操进行降维，并在降维后对数据进行批归一化处理；在后两层卷积层中都使用了步长为2的卷积，并在卷积后同样使用批归一化操作对数据进行归一化处理，最后利用激活函数ReLU的非线性属性对上述结果进行非线性拟合。图2(b)显示了检测较为粗糙的定位结果，导致这一结果的原因大致可归结为以下两个：(1)提供的训练数据让网络模型难以鉴别篡改与非篡改区域之间的图像属性差异；(2)粗筛网络的训练数据大小为$P_{\mathrm{f}}\times P_{\mathrm{f}}$，难以提供足够的信息以区分篡改与非篡改区域之间的图像属性差异。

根据以上所得信息，在粒提网络进行训练之前，为产生已篡改标签的训练数据将沿着篡改图像中篡改区域的边缘取大小为$P_{\mathrm{s}}\times P_{\mathrm{s}}(P_{\mathrm{s}}>P_{\mathrm{f}})$的块，接着为产生未篡改标签的训练数据将在对应的原始图像的边缘上取相同大小的块。然后使用验证集以验证训练好的粗筛网络模型，将验证集中判断错误的块在对应图像中按大小$P_{\mathrm{s}}\times P_{\mathrm{s}}$取出一并作为粒提网络的训练集。构造以上训练集的目的如下：(1)进一步地区分篡改区域边缘和原始图像边缘之间的图像内容差异，减少误检；(2)利用验证集再训练的思想提高网络模型泛化性。最后将所得训练数据输入到粒提网络中进行训练并最终得到检测模型。粒提网络的检测结果(s-out)如图2(c)所示。

2.1.3 精辨网络

根据2.1.2小节中粒提网络的训练思路，本文进一步训练精辨网络。精辨网络的训练数据集构造与粒提网络相似，仅块大小由$P_{\mathrm{s}}\times P_{\mathrm{s}}$扩张到$P_{\mathrm{t}}\times P_{\mathrm{t}}$。精辨网络结构与粒提网络结构相似，仅第1层、第2层卷积层以及最后一层全连接层的神经元个数扩大1倍。精辨网络的检测结果(t-out)如图2(d)所示，可以发现其检测效果较粒提网络有明显提高，可达到进一步地提升检测精度的目的。

2.2 自适应筛选后处理

在经过级联卷积神经网络的检测后，预测得到的可疑篡改区域仍包含一些判定错误的区域。基于此，本文进一步提出自适应筛选后处理来解决以上问题，自适应筛选后处理包含自适应筛选器和填充后处理两个步骤。

首先，为尽可能地填充空隙部分以及滤掉误检部分，对级联卷积神经网络输出的可疑篡改区域t-out使用一个简单的形态学操作并得到过滤结果t-out。之后使用Agglomerative聚类算法[14]将t-out分成n个簇$\{p_1,p_2,\cdots,p_n\}$，并计算每个簇对应的质心$\{c_1,c_2,\cdots,c_n\}$。由质心集合$\{c_1,c_2,\cdots,c_n\}$里的所有元素构成一个外接多边形，同时计算此多边形的几何中心ge。接着计算质心集合$\{c_1,c_2,\cdots,c_n\}$中每个质心到几何中心ge的欧氏距离，得到距离集合$\{d_{1,\mathrm{ge}},d_{2,\mathrm{ge}},\cdots,d_{n,\mathrm{ge}}\}$。

然后，为衡量簇集$\{p_1,p_2,\cdots,p_n\}$的分布情况，通过式 (2)和式 (3)计算出集合$\{d_{1,\mathrm{ge}},d_{2,\mathrm{ge}},\cdots,d_{n,\mathrm{ge}}\}$中所有元素的标准差std (其中$\overline{d_{\mathrm{ge}}}$为集合$\{d_{1,\mathrm{ge}},d_{2,\mathrm{ge}},\cdots,d_{n,\mathrm{ge}}\}$的平均值)，std越小，则表明簇集

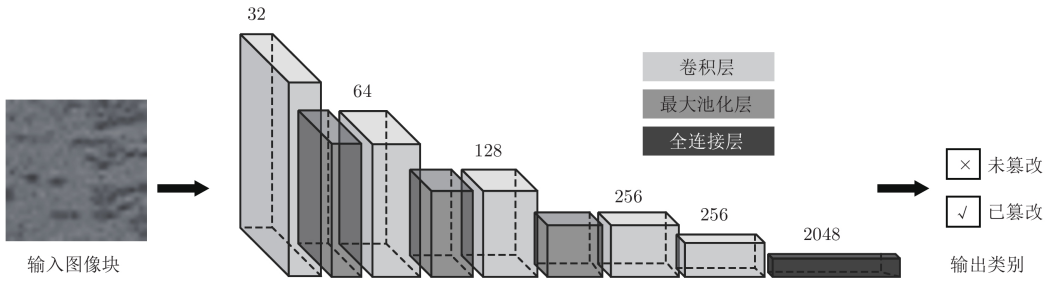


图 4 粒提网络结构

$\{p_1, p_2, \cdots, p_n\}$ 的分布越集中，std越大，则表明簇集$\{p_1, p_2, \cdots, p_n\}$的分布越分散。

$$\overline{d_{\rm ge}} = \frac{1}{n}\sum_{i=1}^{n} d_{i,{\rm ge}} \tag{2}$$

$$\text{std} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{i,{\rm ge}} - \overline{d_{\rm ge}})^2} \tag{3}$$

详细地说，若std小于集合$\{d_{1,{\rm ge}}, d_{2,{\rm ge}}, \cdots, d_{n,{\rm ge}}\}$中任意一个元素，表明簇集$\{p_1, p_2, \cdots, p_n\}$是呈集中分布状态，相应地自适应筛选器的结果$\widetilde{\text{t}-\text{out}}$即等于t-out；若std大于$\{d_{1,{\rm ge}}, d_{2,{\rm ge}}, \cdots, d_{n,{\rm ge}}\}$中某一个或几个元素，表明簇集$\{p_1, p_2, \cdots, p_n\}$呈分散分布状态，则需使用阈值tv去剔除那些分散较远的簇。为计算阈值tv，首先取出包含像素最多的簇p_k，将其设为基准簇并求出对应的质心c_k。然后计算质心集合$\{c_1, c_2, \cdots, c_n\}$中每个质心到质心$c_k$的欧氏距离，并记为$\{D_{1,k}, D_{2,k}, \cdots, D_{k,k}, \cdots, D_{n,k}\}$。通过式(4)即可得到阈值tv

$$\text{tv} = \log_2\left(2n \cdot \prod_{i=1}^{n} D_{i,k}\right) \tag{4}$$

若$D_{i,k}$小于阈值tv，相对应的簇p_i则保留，得到所有保留的簇集即为自适应筛选器的结果$\widetilde{\text{t}-\text{out}}$。最后，使用convex-hull算法[15]对得到的自适应筛选结果$\widetilde{\text{t}-\text{out}}$进行空白填充，并得到最终的检测结果AF-out。

3　实验数据及对比分析

3.1　实验数据

公开数据集CASIA v2.0[1)]是一个图像内容信息较为复杂且更接近于真实数据的图像数据集，其中大多数图像大小接近384×256。为训练级联卷积神经网络并得到最优的检测模型，本文分别针对各级网络来创建数据。本文从CASIA图像数据集中筛选出共计1275组剪切组合篡改图像数据作为实验数据集，从其中随机选取835组图像数据作为粗筛网络、粒提网络和精辨网络共同的训练数据集，并从剩余数据集中随机选取150组图像数据作为粗筛网络的验证数据集，再随机选取150组图像数据作为粒提网络的验证数据集，以及100组图像数据被选作精辨网络的验证数据集，余下的40组图像数据作为测试数据集。

3.1.1　粗筛网络的数据集

为训练粗筛网络鉴别图像中篡改与非篡改区域之间图像属性差异的能力，在训练数据集中，首先沿着篡改图像中篡改区域的边缘取出所有像素点，以每个像素点为中心像素在图像上取出大小为32×32的块($P_{\rm f} = 32$)，并将所得块的标签类别置为已篡改；然后在对应的原始图像上的相同位置取出同样大小的块，并将这些块的标签类别置为未篡改。最终可得到约220000个训练图像块。为验证粗筛网络的鉴别能力，在粗筛网络的验证集中用与粗筛网络产生训练集相同的方式在篡改图像及原始图像上取块，并最终得到约56000个验证图像块。

3.1.2　粒提网络的数据集

为训练粒提网络进一步区分图像中篡改与非篡改区域之间图像属性差异的能力，在训练数据集中，首先沿着篡改图像中篡改区域的边缘取出所有像素点，并将每个像素点作为中心像素在图像上取出大小为64×64的块($P_{\rm s} = 64$)，同时将所得块的标签类别置为已篡改；之后在对应的原始图像上使用Canny检测算法[16]进行边缘检测，并在检测得到的边缘上以100个像素为间隔取出像素点作为中心像素生成同样大小的块，并将这些块的标签类别置为未篡改。然后使用粗筛网络的验证集来验证训练好的粗筛网络模型，将验证集中判断错误的块在对应图像中以大小64×64取出一并作为粒提网络的训练集。最终可得到约200000个训练图像块。以上通过对困难样本特征进行再学习的方式，可进一步地提高网络模型的特征提取及泛化能力。为验证粒提网络的区分能力，在粒提网络的验证集中用与粒提网络产生训练集相同的方式在篡改图像及原始图像上取块，并最终得到约45000个验证图像块。

3.1.3　精辨网络的数据集

为训练和验证精辨网络精准定位图像中篡改区域的能力，与粒提网络中相同的方式产生训练集与验证集，仅图像块大小由64扩张到96($P_{\rm t} = 96$)，最终可得到约100000个训练图像块和30000个验证图像块。

3.2　算法参数及实验评价参数

在级联卷积神经网络的训练过程中，各级网络的初始学习率被设定为0.0001，并选用随机梯度下降作为各级网络的优化器，交叉熵作为各级网络的损失函数。为防止级联卷积神经网络中各级子网络出现过拟合情况，本文使用Dropout[17]以0.8的概率将神经元从网络中暂时丢弃。在自适应筛选后处理过程中，被划分的簇个数为4。级联卷积神经网络实现平台为TensorFlow深度学习库，且实验对比算法都已调至最优参数。本文对比实验部分所参考的评价参数为精确率、召回率和F数，分别如式(5)，式(6)以及式(7)所示

$$\text{精确率} = \frac{\text{真阳性}}{\text{真阳性} + \text{伪阳性}} \tag{5}$$

[1)] CASIA v2.0：<http://forensics.idealtest.org/casiav2/>

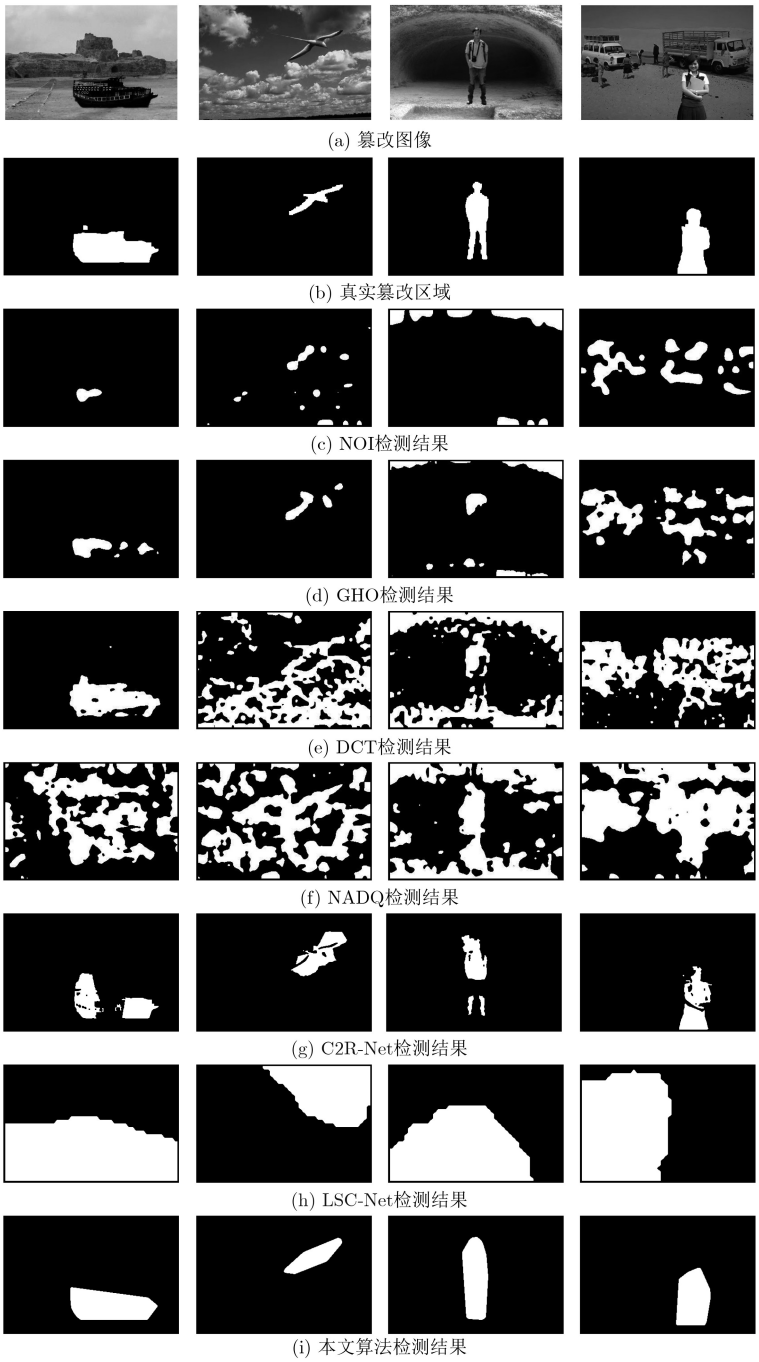
$$召回率 = \frac{真阳性}{真阳性 + 伪阳性} \tag{6}$$

$$F = \frac{2 \times 精确率 \times 召回率}{精确率 + 召回率} \tag{7}$$

其中，真阳性表示检测结果中正确检测为篡改区域的像素总量，伪阳性表示检测结果中错误检测为篡改区域的像素总量，伪阴性表示检测结果中错误检测为非篡改区域的像素总量。

3.3 评价实验及对比分析

为验证本算法的有效性，NOI[1], GHO[2], DCT[3], NADQ[4](以上对比算法由等人复现)及C2R-Net[8], LSC-Net[9]算法被选作对比实验算法。以上6种对比算法中，前4种是传统的基于特征提取的图像篡改检测算法，后两种是当前基于卷积神经网络的图像篡改检测算法。由于GHO, DCT以及NADQ3种算法是基于JPEG图像格式的压缩属性进行检测，而原始CASIA数据集的图像数据格式为TIFF，为保证对比实验的公平性，本文将TIFF格式的图像数据转化为零压缩的JPEG格式的图像数据。从40组图像数据中随机挑选4组数据作为示例，各算法的检测结果如图5所示。图5中每一列表示一组示例及不同方法的检测结果，其中图5(a)



(a) 篡改图像

(b) 真实篡改区域

(c) NOI检测结果

(d) GHO检测结果

(e) DCT检测结果

(f) NADQ检测结果

(g) C2R-Net检测结果

(h) LSC-Net检测结果

(i) 本文算法检测结果

图 5 4组对比实验示例

表示篡改图像，图5(b)表示对应的真实篡改区域图像，图5(c)表示NOI算法的检测结果，图5(d)表示GHO算法的检测结果，图5(e)表示DCT算法的检测结果，图5(f)表示NADQ算法的检测结果，图5(g)表示C2R-Net算法的检测结果，图5(h)表示LSC-Net算法的检测结果，图5(i)表示本文算法的检测结果。

从图5可看出，NOI及NADQ算法对图像进行篡改检测的结果趋于失效状态，GHO算法也仅能大致定位出图像中的篡改内容，C2R-Net可定位篡改区域但存在较多漏检，DCT及LSC-Net的检测结果产生较多误检难以定位到具体区域，而本文算法展示了较好较稳定的检测效果。本文算法及6种对比算法(NOI, GHO, DCT, NADQ, C2R-Net及LSC-Net)各自在40组图像数据上的检测结果的精确率、召回率和F的平均值如表1所示。

从表1可看出，本文算法的检测效果在评价参数精确率及F上都优于其它6种对比算法，但召回率略低于DCT算法。尽管DCT算法的检测结果的召回率较高，但从主观视角来判断，其检测结果存在大范围失效(如图5(e)实例所示)。

为验证本算法的鲁棒性，本文进一步地计算6种对比算法及本文算法各自在40组图像数据面对不同类型、不同程度攻击下的检测结果的精确率、召回率以及F的平均值，检测结果如图6所示。图6中(a1)～(a3)表示6种对比算法及本文算法在不同程度的JPEG图像压缩攻击下的检测结果(例如：当品质因子为100时，即表示无压缩)，图6(b1)～(b3)表示各算法在不同方差(均值默认为0)的高斯噪声污染攻击下的检测结果，图6(a1)和图6(b1)表示各算法在不同类型、不同程度攻击下的检测结果的精确率，图6(a2)和图6(b2)表示各算法在不同类型、不同程度攻击下的检测结果的召回率，图6(a3)和图6(b3)表示各算法在不同类型、不同程度攻击下的检测结果的F。

表 1　各算法的检测结果

算法	精确率	召回率	F
NOI	0.15	0.13	0.14
GHO	0.19	0.27	0.22
DCT	0.42	**0.81**	0.55
NADQ	0.12	0.70	0.20
C2R-Net	0.61	0.43	0.51
LSC-Net	0.14	0.62	0.23
本文算法	**0.62**	0.73	**0.67**

从图6可看出，在不同品质因子的JPEG图像压缩攻击下，NOI, GHO, NADQ3种算法的检测结果趋于失效状态，同时由于CASIA数据集中图像不具有拍摄信息元数据导致LSC-Net算法的检测效果趋于失效，DCT算法在品质因子为100～80范围内具有一定检测效果，但当DCT算法在品质因子为80以下的JPEG图像压缩攻击下时，其检测结果基本失效，C2R-Net算法的检测较为稳定但检测结

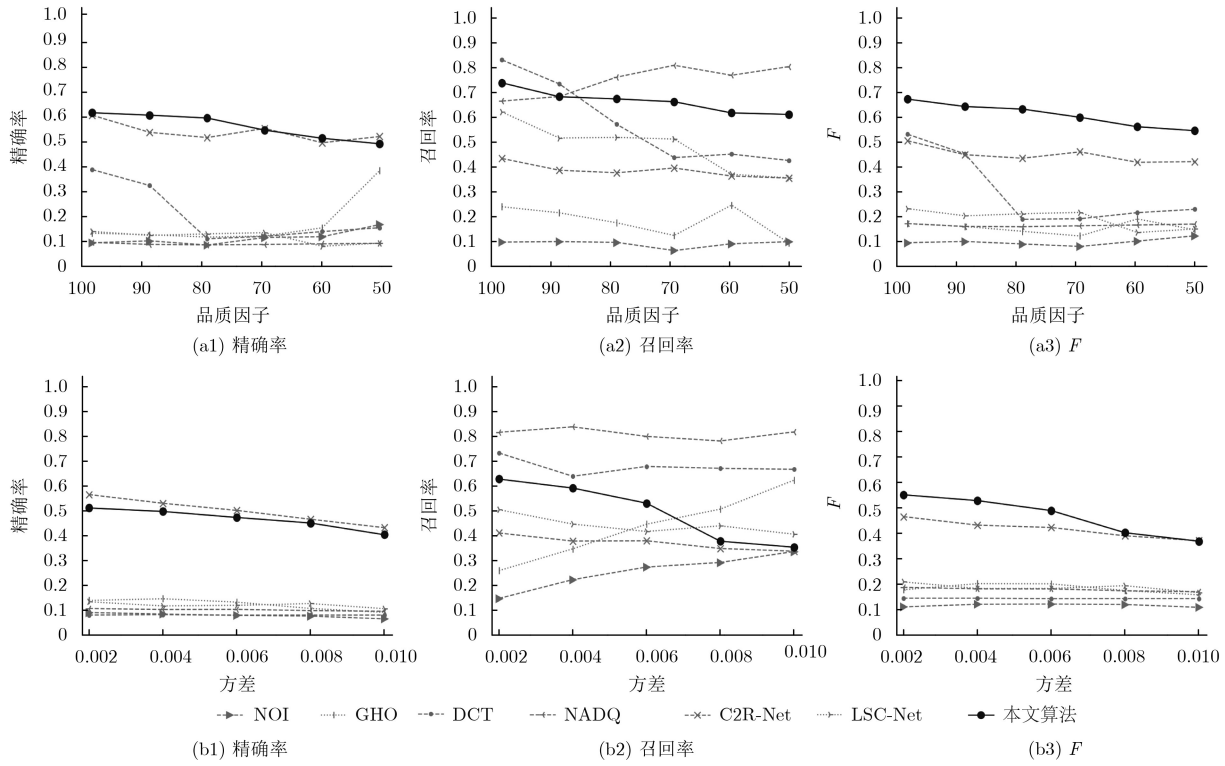


图 6 各算法在不同攻击下的检测效果

果不佳，而本文算法展示了较好较稳定的检测效果；在不同方差的高斯噪声污染攻击下，本文算法的检测效果同样优于6种对比算法，其中NOI, DCT, GHO, NADQ, LSC-Net5种算法的检测结果趋于失效状态。

4 结束语

本文提出了一种基于级联卷积神经网络的图像篡改检测算法。在本文算法中，首先利用级联卷积神经网络对图像中的篡改区域进行定位，接着使用自适应筛选后处理方法进一步对检测结果进行精细优化并得到最终的检测结果。同时，为了验证本算法的有效性及鲁棒性，在数据集CASIA上与当前几种检测效果较好的篡改检测算法进行实验对比，本文算法的最终检测效果优于其它几种方法。

参 考 文 献

[1] MAHDIAN B and SAIC S. Using noise inconsistencies for blind image forensics[J]. *Image and Vision Computing*, 2009, 27(10): 1497–1503. doi: 10.1016/j.imavis.2009.02.001.

[2] FARID H. Exposing digital forgeries from JPEG ghosts[J]. *IEEE Transactions on Information Forensics and Security*, 2009, 4(1): 154–160. doi: 10.1109/TIFS.2008.2012215.

[3] YE Shuiming, SUN Qibin, and CHANG E C. Detecting digital image forgeries by measuring inconsistencies of blocking artifact[C]. 2007 IEEE International Conference on Multimedia and Expo, Beijing, China, 2007: 12–15. doi: 10.1109/ICME.2007.4284574.

[4] BIANCHI T and PIVA A. Image forgery localization via block-grained analysis of JPEG artifacts[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(3): 1003–1017. doi: 10.1109/TIFS.2012.2187516.

[5] RAO Y and NI Jiangqun. A deep learning approach to detection of splicing and copy-move forgeries in images[C]. 2016 IEEE International Workshop on Information Forensics and Security, Abu Dhabi, UAE, 2016: 1–6. doi: 10.1109/WIFS.2016.7823911.

[6] ZHANG Ying, GOH J, WIN L L, *et al*. Image region forgery detection: A Deep Learning Approach[M]. MATHUR A and ROYCHOUDHURY R. Proceedings of the Singapore Cyber-Security Conference. Amsterdam: IOS Press, 2016: 1–11. doi: 10.3233/978-1-61499-617-0-1.

[7] BAPPY J H, ROY-CHOWDHURY A K, BUNK J, *et al*. Exploiting spatial structure for localizing manipulated image regions[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 4980–4989. doi: 10.1109/ICCV.2017.532.

[8] WEI Yang, BI Xiuli, and XIAO Bin. C2R Net: The coarse to refined network for image forgery detection[C]. The 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering, New York, USA, 2018: 1656–1659. doi: 10.1109/TrustCom/BigDataSE.2018.00245.

[9] HUH M, LIU A, OWENS A, *et al*. Fighting fake news: Image splice detection via learned self-consistency[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 106–124. doi: 10.1007/978-3-030-01252-6_7.

[10] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. https://arxiv.org/abs/1409.1556, 2014.

[11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: 10.1109/CVPR.2016.90.

[12] IOFFE S and SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[EB/OL]. https://arxiv.org/abs/1502.03167, 2015.

[13] NAIR V and HINTON G E. Rectified linear units improve restricted boltzmann machines[C]. The 27th International Conference on Machine Learning, Haifa, Israel, 2010: 807-814.

[14] BEEFERMAN D and BERGER A. Agglomerative clustering of a search engine query log[C]. The 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, USA, 2000: 407-416. doi: 10.1145/347090.347176.

[15] BARBER C B, DOBKIN D P, and HUHDANPAA H. The quickhull algorithm for convex hulls[J]. *ACM Transactions on Mathematical Software*, 1996, 22(4): 469–483. doi: 10.1145/235815.235821.

[16] CANNY J. A computational approach to edge detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8(6): 679–698. doi: 10.1109/TPAMI.1986.4767851.

[17] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, *et al*. Dropout: A simple way to prevent neural networks from overfitting[J]. *The Journal of Machine Learning Research*, 2014, 15(1): 1929–1958.

毕秀丽：女，1982年生，副教授，研究方向包括图像处理、多媒体安全和图像取证.

魏 杨：男，1993年生，硕士生，研究方向包括深度学习、图像取证.

肖 斌：男，1982年生，教授，研究方向包括图像处理、模式识别和数字水印.

李伟生：男，1975年生，教授，研究方向包括智能信息处理与模式识别.

马建峰：男，1963年生，教授，研究方向包括计算机网络、信息安全.