·图形图像处理·　　文章编号：1000-3428(2022)04-0213-10　　文献标志码：A　　中图分类号：TP391.41

面向图像篡改取证的多特征融合U形深度网络

路东生，张玉金，党良慧

（上海工程技术大学 电子电气工程学院，上海 201620）

摘　要：随着图像篡改工具的智能化发展，图像篡改不再局限于拼接、移除等某一具体的类型，往往包含多种篡改类型及其组合操作，使得图像篡改取证工作更具挑战性。提出一种端到端的多特征融合U形深度网络，利用编解码网络提取篡改区域与真实区域之间的对比度差异、边缘差异等篡改痕迹，并使用富隐写模型卷积层获取伪造图像的噪声分布不规律信息，从而在无预处理的情况下实现可疑区域的检测并分割出高置信度的篡改区域。在此基础上，使用特征提取模块获取融合的篡改特征，在融合定位模块中利用分级监督策略融合不同分辨率提取的篡改特征，以准确定位篡改区域，实现篡改区域检测与像素级的分割。实验结果表明，基于所提网络的图像篡改取证方法在NIST16和CASIA数据库上的F1值分别为0.841和0.605，与基于MFCN、RGB-N、MANTRA-net等网络的图像篡改取证方法相比，有较优的检测性能和较高的实时性，且对JPEG压缩、缩放等处理具有更强的鲁棒性。

关键词：图像篡改取证；深度神经网络；编解码网络；噪声信息；富隐写模型

开放科学（资源服务）标志码（OSID）：

中文引用格式：路东生，张玉金，党良慧. 面向图像篡改取证的多特征融合U形深度网络[J]. 计算机工程，2022，48(4)：213-222.

英文引用格式：LU D S，ZHANG Y J，DANG L H. Multi-feature fusion U-structure deep network for image tempering forensics[J]. Computer Engineering，2022，48(4)：213-222.

Multi-Feature Fusion U-structure Deep Network for Image Tempering Forensics

LU Dongsheng，ZHANG Yujin，DANG Lianghui

（School of Electronic and Electrical Engineering，Shanghai University of Engineering Science，Shanghai 201620，China）

【Abstract】With the intelligent development of image tampering tools，the types of image tampering are not limited to a specific type such as splicing and remove，but often includes multiple types of tampering and their combined manipulability，making image forensics more challenging. To solve the complex image tampering detection，an end-to-end Multi-Feature Fusion U-Structure deep network for image forgeries detection（MFF-US net）. Using the encoder - decoder architecture to extract the tampering traces，which include the contrast difference，edge and other differences between the manipulated and non-manipulated regions，and using the convolutional layer of the rich models steganalysis to obtain the irregular information of noise distribution on the forgery image，detection of suspicious areas and segmentation of tampered areas with high confidence without pre-processing. On this basis，the feature extraction module is used to obtain the fused tamper features. In the fusion location module，the hierarchical supervision strategy is used to fuse the tamper features extracted from different resolutions，so as to accurately locate the tamper area and realize the tamper area detection and pixel level segmentation. Experiments show the F1 values of the proposed method in the NIST16 and CASIA databases are 0.841 and 0.605，respectively. Compared with the existing MFCN、RGB-N、MANTRA-net mainstream method，this method can not only achieve better detection performance and real-time performance，but also has strong robustness to JPEG compression and post-scaling processing.

【Key words】image tampering forensics；deep neural network；encoder-decoder network；noise information；rich steganography model

DOI：10. 19678/j. issn. 1000-3428. 0061039

基金项目：上海市科委重点项目（18511101600）；上海市自然科学基金（17ZR1411900）。

作者简介：路东生（1996—），男，硕士研究生，主研方向为图像处理、深度学习、图像篡改取证；张玉金（通信作者），副教授、博士；党良慧，硕士研究生。

收稿日期：2021-03-08　　**修回日期：**2021-05-14　　**E-mail：**m020218117@sues.edu.cn

0 概述

智能设备与社交软件的升级迭代，促进了数字图像的应用与发展，主流的图像处理软件如Photoshop、Gimp、美图秀秀等，具有强大的图像编辑功能，让图像篡改操作变得更加便利。数字图像篡改可广泛地分为内容保留与内容改变两类，内容保留包括JPEG压缩[1]、滤波操作[2]、对比度增强等，对图像具有较低的破坏性，并未改变语义信息；内容改变具体分为拼接[3-4]、复制-粘贴[5-6]、移除[7]，这些操作将修改图片内容并导致语义信息改变。复制-粘贴操作在同一张图片中进行，即复制图片中的局部区域并粘贴在同一图片的另一个区域从而形成伪造图片[8-9]，拼接篡改是把来自2张或多张图片中不同的局部区域进行拼接以形成伪造图片，移除篡改是依据图片中的背景区域填补同一图片中被移除的区域。一般来说，改变内容的篡改操作是通过隐藏物体或增加物体数量达到信息误导的目，并结合图像模糊、缩放、扭曲等处理操作使篡改图像检测及定位研究更具挑战性，伪造图像经过专业图像篡改者的加工可以不留下任何视觉线索。

目前，有很多研究工作仅对待检测图像进行分类，即一幅图像是否被篡改，只有少数研究工作尝试进行图像块[10-11]的分类或像素级[12-13]篡改区域定位。相较于图像篡改检测，图像篡改区域的定位同样不可忽视，篡改区域定位能够进一步甄别伪造者的意图，在司法鉴定和法医领域发挥重要作用。此外，多数图像篡改取证方法仅仅关注某一特定的篡改类型，如复制-粘贴、拼接、移除等，但针对单一篡改类型的图像取证方法可能不适用于另一种图像篡改类型，例如由于拼接操作类型来源不同源的图像会引入不同的光电响应、噪声等固有特征，而复制-粘贴操作类型的篡改检测方法不能利用固有特征差异，因此无法对该类型图像进行检测。现实生活中的伪造图像复杂多样，这就要求图像篡改取证研究者的工作不能局限于特定的篡改操作类型。

本文提出一种面向图像篡改取证的多特征融合U形深度网络，以实现端到端的篡改图像检测与定位。利用CNN网络和SRM卷积层提取篡改信息，并将其输入到基于编解码网络和多特征融合的特征提取模块，以实现篡改特征提取。在融合定位模块中利用分级监督策略，结合不同分辨率提取的篡改特征，完成对篡改区域的预测。

1 相关工作

在图像篡改取证研究中，通常根据真实图像与篡改图像间不同特性进行图像检测和篡改区域定位，这些特征包括JPEG压缩效应[1]、边缘不一致[14-15]、噪声模式[16]、色彩一致性、视觉相似度[8-9]、EXIF一致性[3]、相机模型等特性。

待检测图像若曾被复制粘贴，图像中必然存在局部相似的区域，基于此假设，一般的研究方法[7-8]将待检测图像分为非重叠区域和重叠区域，并利用相似性或相关性进行度量，以确定图像块是否被复制，常用的特征提取方法有局部二值模式（Local Binary Patterns，LBP）、方向梯度直方图（Histogram of Oriented Gradient，HOG）、尺度不变特征变换（Scale-Invariant Feature Transform，SIFT）及其改进的算法。文献[5]通过分割待检测图像，对比各个语义独立补丁的仿射变换矩阵以确定匹配点，并进一步匹配确定相似的补丁。文献[17]所提图像块匹配算法能有效用于计算图像上的近似最近邻域，并使用不变特征来匹配相似图像块，例如圆谐波转换，展现了该应用经过几何变换图像块的鲁棒性。在深度学习出现前，研究主要关注判定图像及图像块是否被篡改，由于深度学习在目标检测、语言分割方面取得了优异表现，复制-粘贴取证也有较大进展。文献[7]定义了两分支的神经网络框架，并分别用于提取篡改区域留下的视觉痕迹、区分篡改区域与背景区域，最终实现像素级的检测定位。文献[8]使用卷积神经网络从图像中提取局部块的特征，计算不同块之间的自相关性，并利用点特征提取器定位匹配点，通过反卷积网络定位篡改区域，对于仿射变换、JPEG压缩、模糊等各种已知攻击具有较强的鲁棒性。

若伪造图像经过拼接操作，则拼接区域将引入不同于背景区域的固有特征，例如噪声不连续、篡改区域边缘和色彩不一致等线索。MAHDIAN等[18]利用小波变换原理估计图像块的噪声水平，并设定阈值不断融合领域图像块，根据噪声的局部不一致性进行篡改区域定位。PAN等[19]利用带通滤波器下的峰值浓度与噪声水平的关系检测篡改区域，该方法首先计算每个局部窗口的噪声，接着对这些噪声值进行K-means聚类，最终确定拼接区域。当拼接区域和原始图像内在噪声方差的差异较小时，该方法的检测结果不理想。ZENG等[20]基于主成分分析（Principal Component Analysis，PCA）方法估算每个图像块的协方差矩阵的最小特征值，通过估计较大图像块的噪声水平确定图像块是否为可疑图像块，将较大的图像块继续分割为较小图形块，并再次进行噪声水平估计，该方法能较有效地定位拼接区域。

文献[21]把待检测图像分为水平和垂直的条带，根据局部区域光源颜色的不一致性实现图像块级的拼接区域定位，因深度学习具有高纬数据的特征多级表征学习能力，基于卷积神经网络的方法应运而生。文献[14]使用卷积神经网络提取篡改区域边缘的显著性差异，同时预测篡改区域及其边缘，最终结合几何限制定位篡改区域。文献[4]设计深度稠密匹配层来寻找2个给定图像特征的潜在拼接区

域，并设计了视觉一致性验证模块，该模块通过交叉验证潜在拼接区域上的图像内容来确定检测。文献[3]使用自动记录的照片EXIF元数据作为训练模型的监督信号，以确定图像是否具有自一致性，将自我一致性模型应用于伪造图像的检测和定位。

文献[9]和文献[22]提出基于修复的图像移除取证方法，可以实现无明显痕迹的物体去除[22]。文献[23]提出一种集成的图像移除篡改检测方法，利用中心像素映射加速相似图像对的搜索，减少处理时间的同时维持了较高的精度，然而针对压缩、低通滤波、模糊等攻击伪造图像效果不理想。文献[9]采用2种强化监督策略以引导卷积神经网络（Convolutional Neural Networks，CNN）自动学习修补特征而非图像内容特征，该方法采用编码器-解码器网络结构，以实现在不考虑特征提取的情况下自动检测、去除篡改区域。

文献[24-26]提出针对复合篡改类型的深度学习方法用于篡改取证，文献[12]在生成特征图上使用长短期记忆（Long Short-Term Memory，LSTM）网络建立相邻像素之间的相关性，结合卷积神经网络结构获取篡改区域与背景区域的边界不连续特点，实现端到端的像素级定位。若篡改区域边缘未留下明显痕迹，则篡改性能下降。文献[24]利用LSTM网络捕获篡改引起的重采样特征，同时使用编解码网络结构捕获篡改痕迹，融合特征完成篡改区域的定位。文献[25]在文献[24]基础上充分考虑浅层特征图对篡改定位的影响，采用跳跃连接以避免边缘、纹理等线索的丢失，进一步提升篡改定位精度。文献[26]提出两阶段的篡改方法，先通过复制粘贴检测器判断图像是否经过克隆和移除篡改操作，再结合基于深度学习的重采样检测器判断是否经过拼接和重采样篡改操作，在一定程度上提高了检测性能。

2　本文方法

本文提出面向图像篡改取证的多特征融合U形深度网络（Multi-Feature Fusion U-Structure deep network for image forgeries detection，MFF-US net）用于图像篡改检测和定位，如图1所示（彩色效果见《计算机工程》官网HTML版），该网络包含信息融合、特征提取、区域定位3个模块，相较于现有使用图像分类的预训练模型深度学习方法，MFF-US net是从0开始训练的高效深度学习网络，能够避免过量的参数增加计算量。

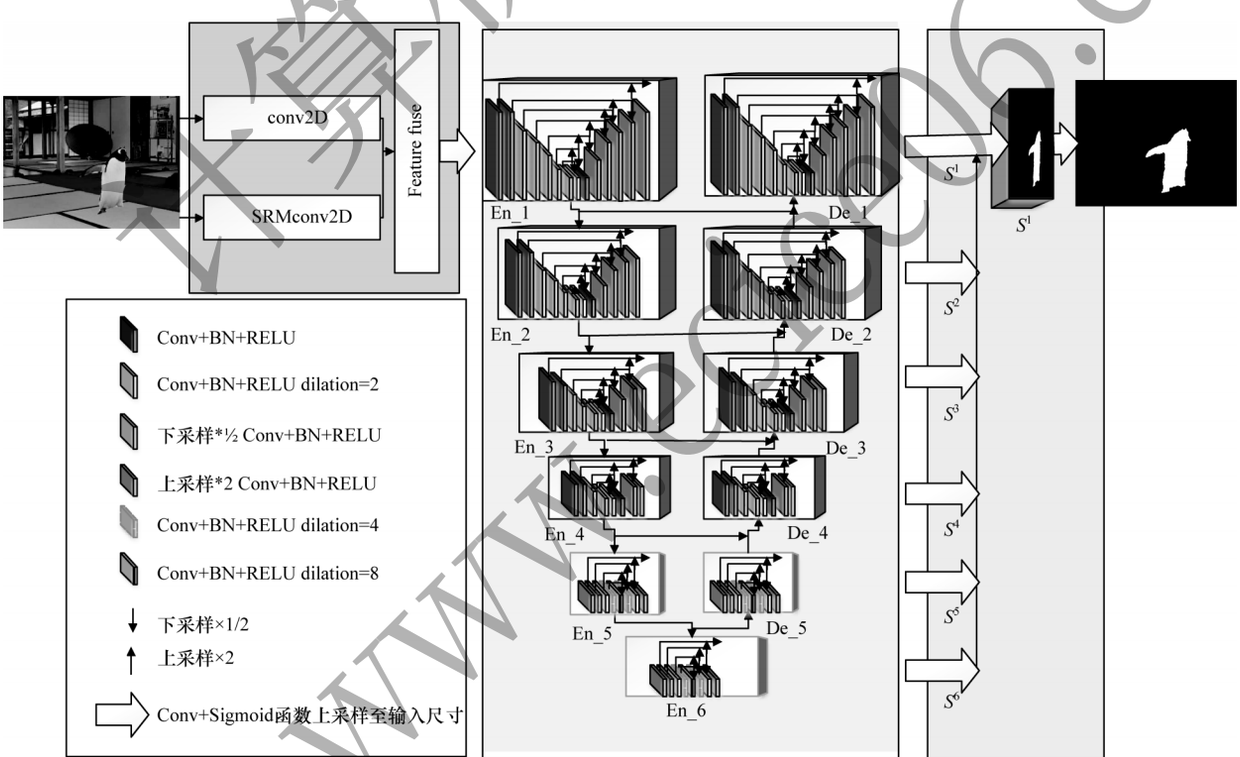


图1　MFF-US net的框架

Fig.1　Framework of MFF-US net

在篡改图像颜色空间，对篡改区域边缘和对比度差异等语义特征建模并不能充分利用篡改区域遗留的噪声痕迹。因此，在信息融合模块中加入富隐写模型卷积层自动提取噪声特征，通过联结操作最大程度地保留篡改线索，并在特征提取模块中利用编码-解码网络多尺度地提取上下文信息。在区域定位模块中，为了避免篡改特征表征的损失，将提取的融合特征分级进行监督并逐层实现特征融合，实现篡改区域检测和高置信度的像素级分类。

本文的贡献主要有以下3个方面：

1)提出一种并不依赖预训练模型的图像篡改取证方法,更加关注篡改区域与真实区域间的特征建模,并在多个公共数据集上取得较优性能。

2)利用SRM模型提取噪声分布特征并融合RGB视觉线索,实现像素级的检测与定位。

3)篡改区域和真实区域存在样本标签不平衡的情况,常篡改区域的像素数量远小于真实的背景区域,因此引入损失函数缓解样本不平衡问题。

3 具体方案

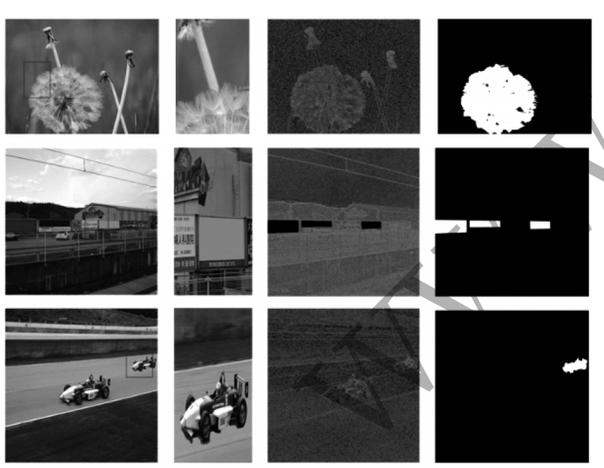
图像篡改取证与目标检测任务相比,后者关注于物体的检测,前者更强调篡改区域遗留的痕迹且要求深度学习网络需要学习更丰富的特征。因此,本文在融合RGB信息和噪声信息的基础上,引入U型残差块[27]构造可堆叠U型结构的MFF-US net,以捕捉更多上下文信息。该网络不同于Hourglss network、Docu-Net、CU-Net等网络[27],其网络的堆叠不会引起计算参数和消耗量被成倍放大,满足高效提取多尺度伪造特征的篡改取证网络。

3.1 信息融合模块

图像作为网络的输入,不需要额外进行预处理操作。在信息融合阶段通过对输入图像进行双分支处理,SRM卷积层和2D卷积层经过卷积处理分别生成相同维度的特征,通过联结所获取的特征作为特征提取模块的输入信息。

3.1.1 RGB信息

复制-粘贴、拼接、移除等图像篡改操作普遍会留下视觉痕迹,在篡改区域形成的过程中,容易造成篡改区域边缘不自然和纹理不连续的现象,如图2(b)所示,篡改区域边缘相较于自然物体边缘更模糊,自然物体边缘视觉上过渡更加自然。



(a)篡改图像 (b)视觉伪影 (c)噪音图像 (d)Ground-truth图像

图2 篡改痕迹示例

Fig.2 Examples of tamper marks

卷积神经网络具有局部感知和参数共享的特点,在目标检测、物体分类等领域表现出较大的潜力,同样能够提取篡改遗留的视觉痕迹。专业的图像篡改者为了使篡改区域与背景区域相似及避免篡改图像上语义信息的不合理性,常使用后处理操作,如旋转、缩放、扭曲、模糊及其组合篡改等操作。从语义信息考虑蒲公英花的拼接、广告牌信息的擦除、赛车数量的增加等符合自然事物的存在,篡改区域边缘的篡改痕迹经过模糊等后处理操作,很难被人察觉,尤其是图2第2行的移除篡改类型,篡改区域为融合背景区域的领域信息,在纹理、对比度等方面无明显差异。经过精心的后处理操作能够使篡改边缘和对比度差异减弱,RGB图像遗留的篡改痕迹并不明显,其采用了噪声特征分支信息来弥补颜色信息空间的不足,因此本文引入局部噪声信息。

3.1.2 噪声信息

相较于RGB图像信息较多关注图像内容提取的低级、高级特征,噪声信息更加注重局部噪声的分布规律。经过篡改的图像必然导致噪声分布不均,图像上的噪声信息作为篡改痕迹的补充,在一定程度上能够解决视觉差异不明显的问题,通过对比噪声估计方法,更好地体现局部噪声特征[28]。采用图3所示的SRM卷积核,图像经过SRM卷积后生成噪声图像,如图2第3列所示。显然,噪声图像强调局部噪声而非图像内容,并能够显示RGB通道中不可见的篡改痕迹,通过相邻元素间的残差建模,形成噪声图表示元素间的共存关系。实验室设置中,2D卷积层的卷积核维度和SRM卷积层相同,维度为$5\times5\times3$,并保证得到相同维度的输出。

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

图3 SRM卷积核

Fig.3 SRM convolution kernel

3.2 特征提取模块

由于现实生活中伪造图像的篡改区域存在大小和形状多样性,因此要求深度学习网络必须具有多尺度特征学习能力,较为常见的处理方法为高频使用1×1、3×3较小卷积核提取特征,以便占用较小的储存空间,以及避免在减少计算量的同时在特征提取阶段只能提取局部特征信息的情况发生。VGG、ResNet、DenseNet[27]等网络并不能满足篡改检测任务对全局信息和局部信息高效提取的要求,他们为提取高分辨率特征图的全局上下文信息,通常使用inception网络,在网络框架的浅层阶段使用空洞卷积增加接受野,但这将导致计算和内存资源消耗增加。为减少计算资源占用,PoolNet在下采样阶段使用较小卷积核代替空洞卷积。由于在多尺度特征融合阶段,上采样和连接操作会导致高分辨率特征信息的损失,因此引入残差U-blocks块作为信息提取的结构。

残差块和残差U型块的结构如图4所示。残差U-blocks能够获取多尺度块内特征,其结构为RSU-L(C_{in},M,C_{out}),如图4(c)所示。

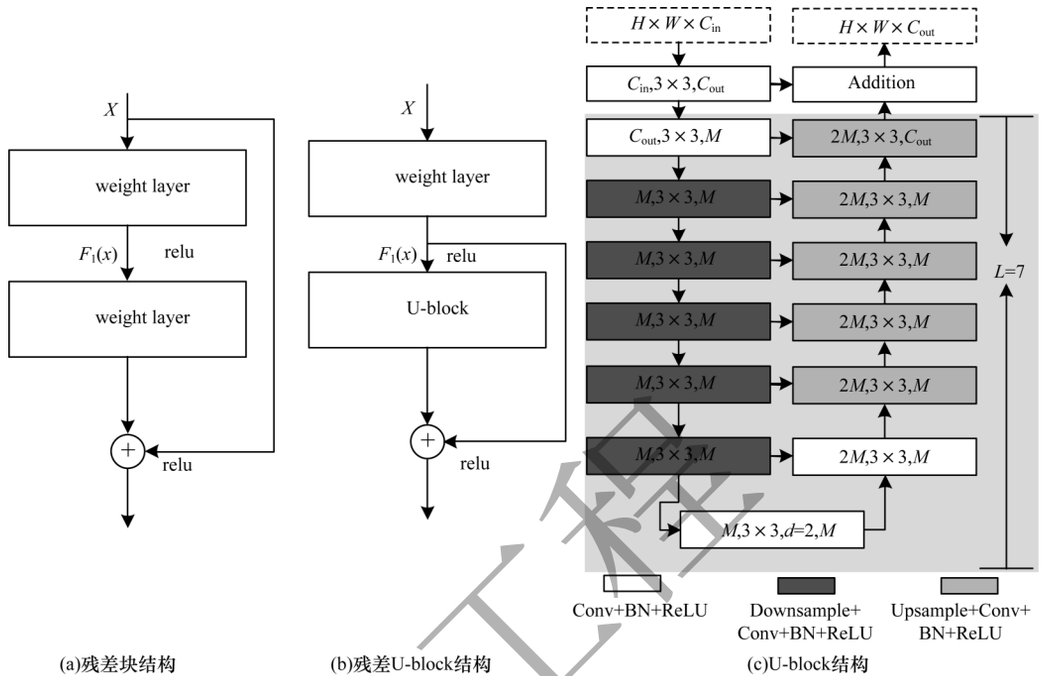


图4　残差块和残差U型块的结构

Fig.4　Structure of the residual block and RSU

在图4(c)中：L代表编码阶段层数；C_{in}、C_{out}分别代表输入、输出特征通道；M为RSU内中间层通道数，主要由3个部分组成：

1)普通卷积层，用于提取局部特征信息，将输入特征图$x(H\times w\times C_{in})$转换为中间映射$F_1(x)$。

2)高度为L的对称U型结构的编码-解码结构，能够提取并编码多尺度上下文信息$u(F_1(x))$，L越大，代表更多的池化操作、更广范围的感受野、更多的局部和全局特征信息。编码-解码结构逐步在下采样的特征图中提取多尺度特征，并通过逐步上采样、拼接和卷积操作编码成高分辨率特征图，这一过程能够减少大尺度特征直接上采样造成的细节损失。

3)残差连接结构，用于融合局部特征和多尺度特征$u(F_1(x))+F_1(x)$。

为更清晰地阐述残差U-blocks和原始残差块的差异，原始残差块被定义为：$H(x)=F_2(F_1(x))+x$，其中：$H(x)$为输入特征x的映射结果；F_2和F_1分别表示权重层的卷积操作。残差U-blocks最大的差异在于使用U型结构代替卷积结构，其定义为：$H_{RSU}(x)=u(F_1(x))+F_1(x)$，其中：$u$代表多尺度的U型框架，由于U型框架较小，在提取多尺度特征的过程中不会消耗明显的计算力。

编码-解码结构是对称结构，能有效提取各分辨率特征图的多尺度信息，避免有效信息的损失。编码阶段如图1所示，共6个块，其中En-1、En-2、En-3、En-4分别为不同L的残差U-block块，相对应的L为7、6、5、4，L取决于输入图像特征图的分辨率大小。对于En-5、En-6块，此时的特征图尺寸较小，进一步的下采样和池化会导致有效信息的丢失。采用空洞卷积代替下采样或者池化操作，此时输入与输出的维度相同。解码结构与编码结构相似，并使用相对应的上采样和拼接操作，能够逐步恢复特征图的分辨率，有效避免特征信息的损失。解码阶段分为5个块，其中，De-1、De-2、De-3、De-4对应编码阶段的残差U-block块，De-5与En-5结构相似。

3.3　区域定位模块

篡改定位融合模块用于生成篡改区域概率图。首先，分别在En-6、De-5、De-4、De-3、De-2和De-1中使用一个3×3的卷积层和一个sigmoid函数，用于产生S^6_{side}、S^5_{side}、S^4_{side}、S^3_{side}、S^2_{side}、S^1_{side}篡改区域概率图。其次，将3×3的卷积层卷积输出(logit)篡改区域映射到输入图像的尺寸，并将他们进行拼接操作，最后通过卷积层和sigmoid函数生成最终篡改区域概率图S^0。

3.4　损失函数

实验设置中，在训练阶段使用深度监督策略，训练损失被定义为：

$$L=\sum_{m=1}^{M}w_{side}^{(m)}l_{side}^{(m)}+w_{fuse}l_{fues} \tag{1}$$

其中：$l_{side}^{(m)}(M=6)$为篡改定位概率图$S_{side}^{(m)}$，$m=1,2,\cdots,6$，l_{fues}为最终的融合篡改定位概率图S^0的损失；$w_{side}^{(m)}$和w_{fuse}分别为每个损失项的权重。传统方法使用标准二值交叉熵计算损失，其定义如式(2)所示：

$$l=-\sum_{(r,c)}^{(H,W)}\left[R_{G(r,C)}\log_a R_{S(r,c)}+\left(1-R_{G(r,C)}\right)\times\log_a\left(1-R_{S(r,c)}\right)\right] \tag{2}$$

其中：(r,c)为像素坐标；(H,W)分别为图片尺寸高和宽；$R_{G(r,C)}$和$R_{S(r,c)}$分别为对应输入图片的Ground-truth和预测出来的概率图。

然而对于篡改检测而言，篡改区域面积与背景区域存在较大差异，易造成不同类间的不均衡。

在式(2)的基础上，增加参数β以控制类别不平衡，定义新的损失函数L_t如式(3)所示：

$$L_t=-\sum_{(r,c)}^{(H,W)}\left[\beta\times R_{G(r,C)}\log_a R_{S(r,c)}+\left(1-\beta\right)\times\left(1-R_{G(r,C)}\right)\log_a\left(1-R_{S(r,c)}\right)\right] \tag{3}$$

其中：$\beta=\frac{|Y-|}{|Y|}$，$1-\beta=\frac{|Y+|}{|Y|}$；$|Y-|$为篡改区域像素数量；$|Y+|$为非篡改区域的像素数量；$|Y|$为图像像素总数。

4 实验

本节将验证本文方法在4个标准公共数据集上的篡改效果，包含NIST Nimble 2016[29](NIST16)、CASIA[30]、COVER[15]以及Columbia dataset[31]数据集，通过F1、AUC、ROC曲线、定位结果等多方面分析模型的泛化能力，同时采用缩放、JPEG压缩等后处理操作实验，分析模型的鲁棒性。

4.1 数据准备

在实验过程中，使用NIST Nimble 2016、CASIA、COVER、Columbia dataset和文献[24]的synthesized数据集共同作为本文实验的训练集。

1)NIST16数据集。应用于竞赛中，包含3种篡改类型，分别为拼接、复制-粘贴和移除，篡改的数字图像经过后处理操作难以通过视觉痕迹观察到，此数据集中的图片具有不同的背景、光照条件和物体，并提供了篡改图像相对应的Ground-truth图像。

2)CASIA数据集。其包含大量物体的复制粘贴和拼接图像，篡改区域经过精心选择、滤波模糊等后处理操作。该数据集提供了相对应的Ground-truth图像，本文使用CASIA 2.0进行训练，在CASIA 1.0中进行测试性能。

3)COVER数据集。专注于复制粘贴的小型数据集，通过缩放、旋转、扭曲、改变光照等手段产生相似的物体形成篡改图像，并利用多种指标衡量篡改图像的相似度，该数据集也提供篡改图像相对应的mask数据。

4)Columbia数据集。含有拼接图像与真实图像共363幅，其中183幅来自不同数码相机拍摄的真实图像，180幅为拼接而成的图像，图像格式为TIFF格式，尺寸大小范围为757×568像素~1 152×768像素，这些图像主要在室内拍摄而成，场景包含走廊、办公桌、人物、盆栽植物等。

本着公平原则，训练数据集的图像数量划分如表1所示。为加强模型的泛化能力，对输入图像进行缩放、随机垂直翻转、裁剪为280×280等操作以避免过拟合现象的出现，图像缩放使用的是双线性插值方法。

表1 不同数据集中训练集和测试集的图像数量划分

Table 1 Image quantity division of training set and test set in different data sets

数据集	训练集	测试集
NIST16	404	160
CASIA	5 123	921
Columbia	135	45
COVER	75	25

4.2 实验参数

在训练模型过程中，采用Pytorch定义深度网络框架，使用单张GPU，利用NVIDIA TITAN RTX GPU在不同设置条件下进行实验，使用Adam优化算法，初始化学习率为0.001，betas=(0.9，0.999)，eps=1×10^{-8}，weight decay=0，通过batch-size为183 000个epoch迭代训练模型。

4.3 实验分析

为定量评价本文方法的有效性，采用F1分数和接收器操作特性曲线(Receiver Operating Characteristic，ROC)作为对比性能的评价标准，F1得分表示对于篡改检测像素水平的评估标准，利用不同的阈值及最高F1得分作为每张图片最终得分，其定义如式(4)所示。正确检测率(True Positive Rate，TPR)和错误检测率(False Positive Rate，FPR)的计算公式如式(5)和式(6)所示，其中，F_{FN}表示篡改像素点被误检测为真实像素点的数量，F_{FP}表示真实像素点被误检测为篡改像素点的数量，T_{TN}表示真实像素点被正确检测出的数量，T_{TP}表示篡改像素点被正确检测出的数量。

$$\mathrm{F1}=\frac{2T_{TP}}{2T_{TP}+F_{FP}+F_{FN}}\times100\% \tag{4}$$

$$\mathrm{TPR}=\frac{T_{TP}}{T_{TP}+F_{FN}}\times100\% \tag{5}$$

$$\mathrm{FPR}=\frac{F_{FP}}{F_{FP}+T_{TN}}\times100\% \tag{6}$$

ROC曲线是描述不同阈值下二分类的预测表现，ROC曲线的面积表示不同方法下二分类的性能表现，其定义为根据不同的分类阈值，即设置判断像素点为篡改像素点的阈值t，若像素点的分类概率≥阈值t(常取t=0.5)，则判定样本为篡改像素点，其中TPR为纵坐标，FPR为横坐标。根据TPR和FPR的值不同，将他们的值绘制形成曲线，即为ROC曲线。

4.3.1 与现有方法的对比

现有的图像篡改取证方法分为传统的手工特征提取网络和基于深度学习的篡改网络，本节对比现有方法，采用消融实验验证本文方法的有效性，实验中采用以下方法进行对比。

1)ELA[32]方法，通过查找在不同压缩因子情况下篡改区域与背景区域间的压缩错误差异以定位篡改区域。

2）NOI1方法，利用高频的小波系数来模拟局部噪声，设定阈值并不断融合领域图像块，依据噪声的局部不连续性进行篡改区域定位[18]。

3）CFA1方法，假设图像是使用一个彩色滤波器阵列获得的，并且篡改消除了由马赛克算法产生的伪影，通过在局部水平上测量CFA伪影的存在推理出篡改区域[33]。

4）MFCN方法，构造全卷积网络实现篡改边缘和初步篡改区域的预测，利用几何知识整合篡改边缘和初步篡改区域并确定最终篡改区域[14]。

5）J-LSTM方法，联合LSTM网络和CNN网络完成篡改块的判定和像素级的篡改区域分割[12]。

6）RGB-N方法，通过利用双线性池化融合图像信息和噪声信息实现篡改区域的定位[34]。

7）MANTRA-NET方法，利用CNN网络解决篡改痕迹提取和局部异常检测问题，实现篡改区域像素级的定位[35]。

8）Single-RGB方法，本文所提方法采用单流输入的方式，即只考虑RGB信息的输入，记为Single-RGB。

9）Single-Noise方法，Single-RGB，本文所提方法采用单流输入的方式，即只考虑噪声信息的输入，记为Single- Noise。

对比现有方法包括ELA、NOI、CFA1、MFCN、RGB-N和本文方法的F1指标，结果如表2所示。其中：对比Single-RGB、Single-Noise和本文方法可知，具有融合特征的网络优于单流输入的噪声信息和RGB信息。在NIST16、COVER和CASIA数据集上的数据结果可知，Single-RGB方法略优于Single-Nois。然而在Columbia数据集中，Single-Noise取得的F1值比Single-RGB方法高2.3个百分点，原因是Columbia为未压缩的拼接图像，噪声差异较为明显，并未受到后处理操作的影响。

表2　不同方法在不同数据集上的F1值对比

Table 2　Comparison of F1 values of different methods on different data sets

方法	NIST16数据集	Columbia数据集	COVER数据集	CASIA数据集
ELA方法	0.236	0.470	0.222	0.214
NOI方法	0.285	0.574	0.269	0.263
CFA1方法	0.174	0.467	0.190	0.207
MFCN方法	0.571	0.612	—	0.541
RGB-N方法	0.722	0.697	0.437	0.408
Single-RGB方法	0.804	0.688	0.394	0.550
Single-Noise方法	0.793	0.710	0.353	0.512
本文方法	0.841	0.723	0.418	0.605

基于深度学习的篡改检测方法要远优于传统特征提取方法，单一特征的篡改取证方法容易导致多数伪造图像的检测任务失败，这是因为ELA、NOI、CFA1特征提取方法只强调单一的篡改痕迹，且多类型的篡改取证需要更丰富的区分特征。

本文方法在NIST16、Columbia、CASIA数据集上表现较优，分别高于RGB-N方法11.9、2.6和19.7个百分点。在深度学习方法中，MFCN方法的表现性能较差，这是因为在特征提取过程中，采用较小尺寸的卷积核和上采样操作容易致使低层特征损失及较小篡改区域检测效果不理想。与RGB-N方法相比，本文方法采用了RSU结构和分级监督策略，具有丰富的多尺度特征，在一定程度上能够避免较大篡改区域的边缘与较小篡改区域的细节丢失。由表2还可知，本文方法在COVER数据集上的F1值低于RGB-N方法1.9个百分点，这是因为复制粘贴操作产生类似的噪声分布不利于产生区分特征。由此可见本文所提方法的综合性能优于现有方法。

4.3.2　ROC曲线

本节采用ROC曲线对比不同方法的性能，包括ELA、NOI、CFA1、J-LSTM、MANTRA-NET和RGB-N方法，其中ROC曲线与横轴坐标轴形成的区域面积称为AUC值，AUC值越高代表该方法的泛化能力越强。

如表3所示，与基于CNN的深度学习方法相比，ELA方法、NOI方法和CFA1方法因无法实现通用的取证模型表现出较弱的泛化能力，通过对比本文方法、Single-RGB方法和Single-Noise方法在不同数据集上的AUC值高低，验证了本文方法的有效性。其中，J-LSTM方法利用CNN提取浅层特征图并分块输入LSTM网络中，在一定程度上造成篡改区域的边缘定位不准确，在NIST16和COVER数据集上的AUC值分别为0.764和0.614，泛化能力较弱。本文方法在NIST16、Columbia、CASIA数据集上的AUC值均为最高，分别高于MANTRA-NET方法14.7、8.5和2.8个百分点，且MANTRA-NET方法利用多层CNN提取特征过程中易造成浅层特征的丢失，如篡改区域的边缘等细节不准确。本文所提方法在NIST16、Columbia、COVER和CASIA数据集上的AUC值分别为0.942、0.909、0.727和0.845，其相对应的ROC曲线为图5所示像素级分割的ROC曲线，由图5可知由于不同数据集分布不同，单一的阈值设置并不能取得最优性能，这进一步说明了本文所提方法具有较强的泛化能力。

表3　在标准数据集上AUC值的比较

Table 3　Comparison of AUC values on the standard data set

方法	NIST16数据集	Columbia数据集	COVER数据集	CASIA数据集
ELA方法	0.429	0.581	0.583	0.613
NOI方法	0.487	0.546	0.587	0.612
CFA1方法	0.501	0.720	0.485	0.522
J-LSTM方法	0.764	—	0.614	—
MANTRA-NET方法	0.795	0.824	0.819	0.817
RGB-N方法	0.937	0.858	0.817	0.795
Single-RGB方法	0.852	0.807	0.658	0.845
Single-Noise方法	0.856	0.886	0.543	0.807
本文方法	0.942	0.909	0.727	0.845

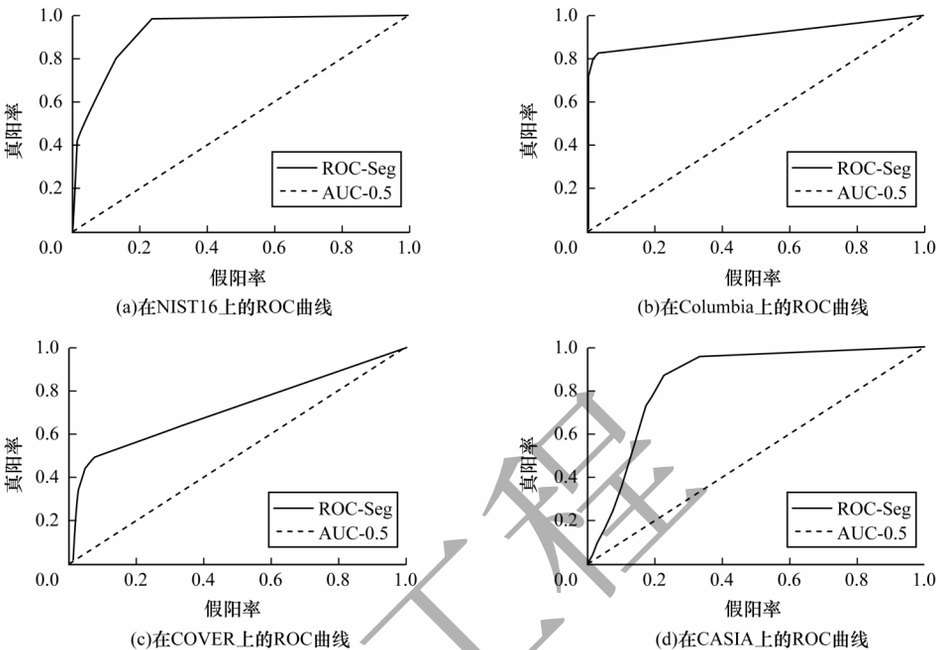


图5 在4个标准数据集上像素级分类的ROC曲线

Fig.5 ROC curve of pixel-level classification on four standard data sets

4.3.3 定位结果分析

为进一步验证本文方法的有效性,本节对一些伪造图像进行了篡改检测与定位,图6所示的是来源于4个标准数据集中的篡改检测实例,其中包括篡改图像、噪声图像、Ground-truth图像以及本文方法的检测结果。图6(a)、图6(b)、图6(c)和图6(d)分别来自数据集NIST16、Columbia、COVER和CASIA,包括拼接、复制-粘贴和移除篡改类型,第1列为待检测的篡改图像直接输入到网络模型中,无需缩放等预处理操作,第2列为噪声图像,第3列为Ground-truth图像,第4列为本文方法的输出结果。显然,由于MFF-US网络具有多尺度高分辨率特征提取能力,篡改区域能够应对任意图像尺寸的篡改检测,并且在较小篡改区域检测和较大篡改区域边缘均取得高置信度的检测结果。从图6(c)和图6(d)中可以发现检测结果存在漏检篡改区域的情况,对于多数篡改图像能够精确地检测并分割篡改区域。

4.3.4 鲁棒性分析

JPEG图像压缩及几何变换是常见的拼接图像后处理操作,为进一步评估本文所提方法的鲁棒性,统计NIST16数据库中的测试集分别经过压缩因子QF=70的JPEG压缩,QF=50的JPEG压缩缩放0.7和缩放0.5操作后检测的F1值,结果如表4所示。由表4可知,本文方法相较于现有其他方法具有较强的抗缩放和抗JPEG攻击的能力,在压缩因子为70和缩放0.7的情况下,F1值略有降低,分别减少2.3和2.7个百分点,在压缩因子为50和缩放0.5的情况下,F1值有明显下降,分别减少10.3和5.4个百分点,本文方法的F1值相较于RGB-N方法分别提高了6.1和10.6个百分点。综上所述,本文所提方法具有较强的鲁棒性和泛化能力。

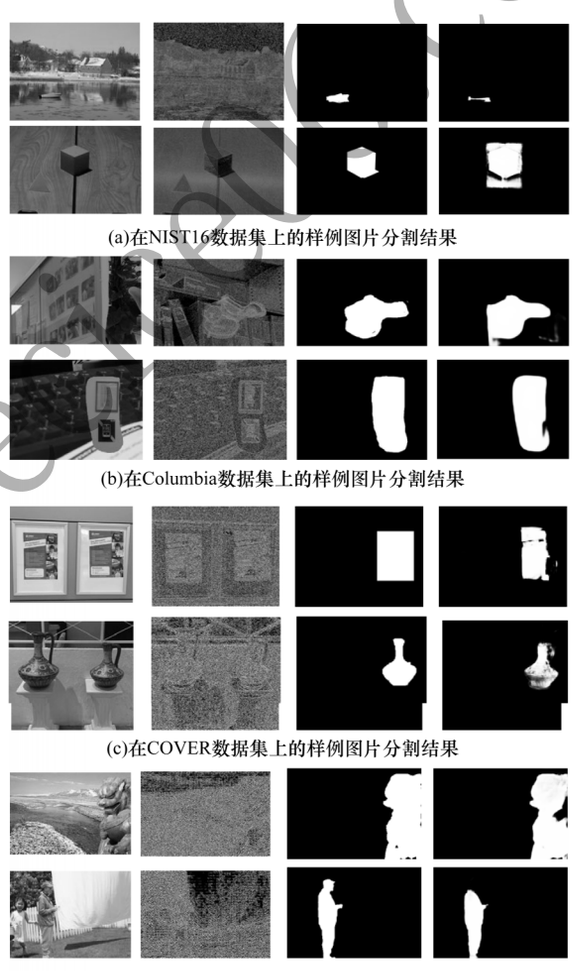


图6 不同数据集的篡改检测结果示例

Fig.6 Examples of tamper detection results for different data sets

表4　不同方法在NIST16测试集JPEG压缩和缩放情况下的F1值

Table 4　F1 value of different methods under JPEG compression and scaling of NIST16 test set

方法	压缩因子为100/缩放1.0	压缩因子为70/缩放0.7	压缩因子为50/缩放0.5
ELA方法	0.285/0.285	0.142/0.147	0.140/0.155
NOI方法	0.236/0.236	0.119/0.141	0.114/0.114
CFA1方法	0.174/0.174	0.152/0.134	0.139/0.141
RGB-N方法	0.722/0.722	0.677/0.689	0.677/0.681
本文方法	0.841/0.841	0.828/0.814	0.738/0.787

4.3.5　复杂度分析

本节针对现有基于深度学习的方法复杂度进行分析，结果如表5所示。表5所示为不同方法模型的参数量和图像的平均推理帧率，用于图像伪造取证的本文方法参数量为168M，仅高于MANTRA-net参数量，远低于MFCN和RGB-N模型参数量，这是因为采用残差U-blocks结构代替常用的卷积层有助于减少模型空间和时间复杂度。在时间复杂度方面，本文方法的帧率达20 frame/s，高于其他方法，能有效满足现实生活中对于篡改取证实时性和有效性的需求。

表5　不同方法模型参数量和耗时的对比

Table 5　Comparison of model parameters and time-consuming of different methods

方法	参数量/10^6	帧率/(frame·s^{-1})
MFCN方法	512.92	—
MANTRA-NET方法	72.80	1.25
RGB-N方法	1 181.82	0.35
本文方法	168.00	20.00

5　结束语

本文提出一种用于图像伪造取证的高效U形深度网络MFF-US net，实现篡改区域的检测与分割。利用CNN网络和SRM卷积层构建特征融合模块，以提取并融合RGB和噪声信息。同时，引入RSU结构并构造出具有多尺度特征的噪声提取模块，并在融合定位模块利用分级监督策略，以融合不同分辨率提取的篡改特征，实现篡改区域检测与像素级的分割。实验结果表明，基于编解码网络和多特征融合的取证方法能够自动学习篡改特征，且无需考虑特征提取和分类设计。与MFCN、RGB-N、MANTRA-net等现有方法相比，本文方法在多个标准篡改取证数据集上均取得较优性能，针对缩放、JPEG压缩等攻击操作具有较强的鲁棒性。下一步将通过生成对抗网络，产生更丰富的篡改数据，加强篡改取证中小目标检测，以应对复杂伪造图像的情况。

参考文献

[1] AMERINI I, URICCHIO T, BALLAN L, et al. Localization of JPEG double compression through multi-domain convolutional neural networks[EB/OL]. [2021-02-02] https://arxiv.org/abs/1706.01788.

[2] CHEN J, KANG X, LIU Y, et al. Median filtering forensics based on convolutional neural networks[J]. IEEE Signal Processing Letters, 2015, 22(11): 1849-1853.

[3] HUH M, LIU A, OWENS A, et al. Fighting fake news: image splice detection via learned self-consistency[C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 101-117.

[4] WU Y, ABD-ALMAGEED W, NATARAJAN P. Deep matching and validation network: an end-to-end solution to constrained image splicing localization and detection[C]// Proceedings of the 25th ACM International Conference on Multimedia. New York, USA: ACM Press, 2017: 1480-1502.

[5] COZZOLINO D, POGGI G, VERDOLIVA L. Efficient dense-field copy-move forgery detection [J]. IEEE Transactions on Information Forensics and Security, 2015, 10(11): 2284-2297.

[6] 邢文博，杜志淳. 数字图像复制粘贴篡改取证[J]. 计算机科学，2019，46(S1)：380-384，396.
XING W B, DU Z C. Digital image forensics for copy and paste tampering[J]. Computer Science, 2019, 46(S1): 380-384, 396. (in Chinese)

[7] 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.

[8] WU Y, ABD-ALMAGEED W, NATARAJAN P. Busternet: detecting copy-move image forgery with source/target localization[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 168-184.

[9] WU Y, ABD-ALMAGEED W, NATARAJAN P. Image copy-move forgery detection via an end-to-end deep neural network[C]//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision. Washington D. C., USA: IEEE Press, 2018: 1907-1915.

[10] 霍占强，刘玉洁，付苗苗，等. 基于卷积神经网络的直线描述方法研究[J]. 计算机工程，2021，47(5)：251-259.
HUO Z Q, LIU Y J, FU M M, et al. Research on line description method based on convolutional neural network[J]. Computer Engineering, 2021, 47(5): 251-259. (in Chinese)

[11] BUNK J, BAPPY J H, MOHAMMED T M, et al. Detection and localization of image forgeries using resampling features and deep learning[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Press, 2017: 1881-1889.

[12] BAPPY J H, ROY-CHOWDHURY A K, BUNK J, et al. Exploiting spatial structure for localizing manipulated image regions [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 4970-4979.

[13] BONDI L, LAMERI S, GUERA D, et al. Tampering detection and localization through clustering of camera-based CNN features[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Press, 2017: 1855-1864.

[14] SALLOUM R, REN Y, KUO C C J. Image splicing localization using a multi-task fully convolutional network[J]. Journal of Visual Communication and Image Representation, 2018, 51: 201-209.
[15] ZHANG Z, ZHANG Y, ZHOU Z, et al. Boundary-based image forgery detection by fast shallow cnn[C]//Proceedings of the 24th International Conference on Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 2658-2663.
[16] COZZOLINO D, VERDOLIVA L. Noiseprint: a CNN-based camera model fingerprint[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 144-159.
[17] LI J, LI X, YANG B, et al. Segmentation-based image copy-move forgery detection scheme[J]. IEEE transactions on information forensics and security, 2014, 10(3): 507-518.
[18] MAHDIAN B, SAIC S. Using noise inconsistencies for blind image forensics[J]. Image and Vision Computing, 2009, 27(10): 1497-1503.
[19] PAN X, ZHANG X, LYU S. Exposing image splicing with inconsistent local noise variances[C]//Proceedings of 2012 IEEE International Conference on Computational Photography. Washington D. C., USA: IEEE Press, 2012: 1-10.
[20] ZENG H, ZHAN Y F, KANG X G, et al. Image splicing localization using PCA-based noise level estimation[J]. Multimedia Tools and Applications, 2017, 76(4): 4783-4799.
[21] FAN Y, CARRÉ P, FERNANDEZ-MALOIGNE C. Image splicing detection with local illumination estimation[C]// Proceedings of 2015 IEEE International Conference on Image Processing. Washington D. C., USA: IEEE Press, 2015: 2940-2944.
[22] 朱新山, 钱永军, 孙彪, 等. 基于深度神经网络的图像修复取证算法[J]. 光学学报, 2018, 38(11): 105-113.
ZHU X S, QIAN Y J, SUN B, et al. Image inpainting forensics algorithm based on deep neural network[J]. Acta Optica Sinica, 2018, 38(11): 105-113. (in Chinese)
[23] 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.
[24] 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.
[25] MAZAHERI G, MITHUN N C, BAPPY J H, et al. A skip connection architecture for localization of image manipulations [EB/OL]. [2021-02-02]. https://www. researchgate. net/publication/335463706_A_Skip_Connection_Architecture_for_Localization_of_Image_Manipulations.
[26] MOHAMMED T M, BUNK J, NATARAJ L, et al. Boosting image forgery detection using resampling features and copy-move analysis[J]. Electronic Imaging, 2018, 48(7): 118-128.
[27] QIN X, ZHANG Z, HUANG C, et al. U2-Net: going deeper with nested U-structure for salient object detection[EB/OL]. [2021-02-02]. https://arxiv. org/abs/2005. 09007.
[28] FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.
[29] MULTIMODAL INFORMATION GROUP. Open media forensics challenge[EB/OL]. [2021-02-02]. https://www.nist. gov/itl/iad/mig/open-media-forensics-challenge.
[30] JIE Z, WEI F Z. CASIA tampered image detection evaluation dataset[EB/OL]. [2021-02-02]. https://www. oalib. com/references/14079387.
[31] TIAN T N, JESSIE H, SHIH F C, Columbia Image splicing detection evaluation dataset[EB/OL]. [2021-02-02]. https://www. ee. columbia. edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet. htm.
[32] KRAWETZ N, SOLUTIONS H F. A picture's worth[J]. Hacker Factor Solutions, 2007, 6(2): 2-6.
[33] FERRARA P, BIANCHI T, DE ROSA A, et al. Image forgery localization via fine-grained analysis of CFA artifacts[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(5): 1566-1577.
[34] ZHOU P, HAN X, MORARIU V I, et al. Learning rich features for image manipulation detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 1053-1061.
[35] WU Y, ABDALMAGEED W, NATARAJAN P. Mantra-net: manipulation tracing network for detection and localization of image forgeries with anomalous features[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 9543-9552.

编辑 赖玉玲