文章编号：1003-0530(2021)12-2302-21

图像源辨识取证研究综述

陈艺芳[1,2] 何自强[2] 文冠臣[1] 康显桂[2]
(1. 广东技术师范大学网络空间安全学院，广东广州 510665；2. 中山大学广东省信息安全技术重点实验室，广东广州 510006)

摘 要：图像源辨识取证是数字图像被动取证的一个重要研究方向，旨在从图像本身的特性出发对图像的相关来源进行识别和检测。图像源辨识取证主要包括：相机源识别、计算机图形学方法生成图像取证、AI 合成图像的取证以及重获取图像的取证。其中相机源识别主要包括对采集图像所用相机的品牌、型号或个体进行识别。计算机图形学方法生成图像取证、AI 合成图像的取证以及重获取图像的取证都属于非自然图像(包括生成、合成、重获取图像)与自然图像间的溯源分析。近年来，在图像源辨识取证方面的研究已经取得了一些研究成果。现有的研究方法大致分为两类，包括传统的基于模型的方法和基于深度学习的方法。本文整理了图像源辨识取证领域的研究方法，主要对研究框架和基本思路以及常用的评价指标、数据集进行了介绍，最后总结了当前图像源辨识取证的研究现状。

关键词：数字图像取证；图像源辨识取证；相机源识别；图像溯源分析
中图分类号：TP391 **文献标识码：**A **DOI：**10.16798/j.issn.1003-0530.2021.12.005

引用格式：陈艺芳，何自强，文冠臣，等. 图像源辨识取证研究综述[J]. 信号处理，2021，37(12)：2302-2322. DOI：10.16798/j.issn.1003-0530.2021.12.005.

Reference format：CHEN Yifang，HE Ziqiang，WEN Guanchen，et al. A survey of forensics research on image source identification forensics[J]. Journal of Signal Processing，2021，37(12)：2302-2322. DOI：10.16798/j.issn.1003-0530.2021.12.005.

A Survey of Forensics Research on Image Source Identification Forensics

CHEN Yifang[1,2] HE Ziqiang[2] WEN Guanchen[1] KANG Xiangui[2]
(1. College of Cyberspace Security，Guangdong Polytechnic Normal University，Guangzhou，Guangdong 510665，China；
2. Guangdong Key Laboratory of Information Security，Sun Yat-sen University，Guangzhou，Guangdong 510006，China)

Abstract：Image source identification forensics is an important research direction of passive forensics of digital image，which aims to identify and detect the relevant sources of images based on the characteristics of the images themselves. Image source identification forensics mainly includes camera source identification，forensics of computer-generated (CG) image，forensics of AI synthesized images and forensics of recaptured image. Camera source identification mainly includes the identification of the brand，model or individual of the camera used for capturing image. Forensics of computer-generated (CG) image，forensics of AI synthesized images and forensics of recaptured image are source analysis between unnatural images (including generated，synthesized and reacquired images) and natural images. In recent years，some achievements have been made in the field of image source identification forensics. Existing research methods can be roughly divided into two categories，including traditional model-based methods and deep learning-based methods. In this paper，the research methods in the field of image source identification forensics were summarized. The research framework，basic ideas，commonly used evaluation indicators and data sets were introduced. Finally，the current research status of image source identification forensics were summarized.

Key words：digital image forensics；image source identification forensics；camera source identification；image source analysis

收稿日期：2021-07-04；修回日期：2021-09-19
基金项目：国家自然科学基金项目(62102100，61772571，62072484)；国家重点研发计划课题(2019QY2203)

1　引言

随着人工智能和大数据时代的来临，数字图像的广泛传播以及图像处理工具的泛滥，图像被修改和编辑的代价大大降低。互联网的广泛普及使得人们可以轻松便捷地获取功能强大、操作简单的图像编辑软件（如 Adobe Photoshop，ACDSee，美图秀秀等）。另一方面，以生成对抗网络（Generative Adversarial Network，GAN）为代表的基于人工智能（Artificial Intelligence，AI）技术的图像合成工具快速发展，借助这些图像合成工具，即使是没有专业知识背景的用户也可以轻松地生成逼真度高的伪造图像。图像编辑和合成工具的发展对图像内容真实性提出了严重的挑战，“眼见未必为实”的时代已经到来。恶意伪造、以假乱真的图像可能成为事实证据用于法庭举证、新闻报道、政府或重要团体的公告声明等场合，其所导致的误判、误报道和欺诈等问题可能会带来巨大的社会、政治和舆论风险。伪造或篡改的数字图像在各种网络场景下尤其是在社交网络上便捷广泛的传输，这给国家和社会稳定带来更多的安全隐患。

为了更好的保障社会公共秩序，维护司法公正、新闻诚信和网络内容安全，为互联网内容管控和过滤提供技术支持，及时充分地开展针对数字图像取证的研究，是信息安全领域迫切的现实需求。通过数字图像取证技术可以检测和揭露伪造图像，从而保证数字图像的真实性、完整性、可靠性。数字图像取证包括主动取证和被动取证[1]。主动取证技术通过在原始图像中嵌入特定的认证信息来保护图像的完整性和真实性。被动取证也称为“盲取证”，只需根据图像本身的特性就能实现对图像真实性、完整性和可靠性的验证，不需要对数字图像进行认证信息嵌入等预处理操作，具有更强的实用性[2-5]。

图像源辨识取证是数字图像被动取证的一个重要研究方向。本文聚焦图像源辨识取证，从相机源识别、计算机图形学方法生成图像取证、AI 合成图像的取证和重获取图像的取证四个方面分别进行介绍，涵盖了近年来图像源辨识取证领域大部分的研究报道。本文第 2 部分首先对图像源辨识取证的分类和研究方法做了概述。随后，本文在第 3、4、5、6 部分对相机源识别、计算机图形学方法生成图像取证、AI 合成图像取证和重获取图像取证的研究分别进行了详细地总结和叙述。本文第 7 部分介绍了图像源辨识取证中常用的评价指标和数据库。最后在第 8 部分，本文分析总结了当前研究所面临的一些问题，以期推动图像源辨识取证领域研究工作的进一步发展。

2　图像源辨识取证概述

图像源辨识取证是指从图像本身的特性出发，对图像的相关来源进行识别和检测。不同设备拍摄或者不同生成方式产生的数字图像在视觉效果上没有明显的差异，但这些不同来源的图像会具有一些不同的统计特征。现有的图像源辨识取证方法就是通过分析和提取这些能够区别图像来源的特征，对数字图像的来源进行盲取证。图像源辨识取证的分类如图 1 所示，主要包括：相机源识别、计算机图形学方法生成图像取证、AI 合成图像取证和重获取图像取证。相机拍摄的、没有经过任何图像编辑和篡改操作的照片属于自然图像。相机源识别是研究如何从自然图像中提取与拍摄设备特性相关的痕迹，从而辨别拍摄相机的品牌、型号或个体。计算机图形学方法生成图像取证、AI 合成图像的取证以及重获取图像的取证属于非自然图像与自然图像间的溯源分析，即判断图像是由其他生成

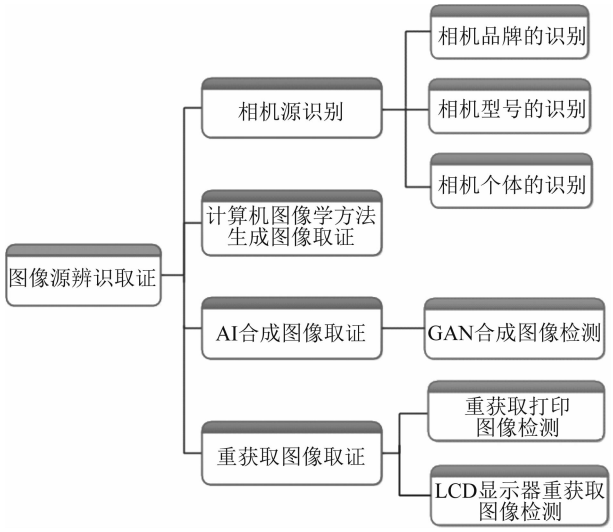

图 1　图像源辨识取证的分类

Fig. 1　The classification of image source identification forensics

方式产生的非自然图像(主要包括生成/合成/重获取图像)还是相机拍摄而成的自然图像。计算机图形学方法生成图像取证的目的是检测图像是否为利用计算机图形学方法生成的图像。AI 合成图像的取证是用于检测图像是否为通过 AI 技术合成的图像,主要包括对生成对抗网络(Generative Adversarial Network, GAN)方法合成图像的检测。重获取图像的取证是检测图像是否为经过了一次中间媒介映射、再利用图像采集设备重新获取得到的图像,主要包括对重获取打印图像和 LCD 显示器重获取图像的检测。

近年来,在图像源辨识取证方面的研究已经取得了一些研究成果,研究方法大致分为两类,包括:传统的基于模型的方法和基于深度学习的方法。基于模型的图像取证方法的框架图如图 2 所示。主要包括训练过程和测试过程两部分。在训练过程中,主要依靠人工经验,手动设计与取证相关的特征,取证框架通常借鉴模式分类的方法,包括特征设计、特征提取,以及训练分类器、获得取证模型这几个步骤。测试过程首先提取特征,再利用训练好的取证模型对未知图像进行分类检测。传统的基于模型的图像源辨识取证方法主要还是依靠人工经验构造源辨识取证特征,特征性能受到设计者先验知识的限制;另外由于特征提取和分类器设计是两个独立的过程,不能同时优化,因此分类性能(如分类精度、分类速度、通用性等)受到一定限制。目前已有越来越多研究者提出了基于深度学习的图像源辨识方法。深度学习用于图像源辨识,实现了直接对图像数据进行自动、逐层的特征学习,并将特征提取和分类器作为一个整体进行优化,从而提高了图像源辨识特征的表达能力和分类器的性能。

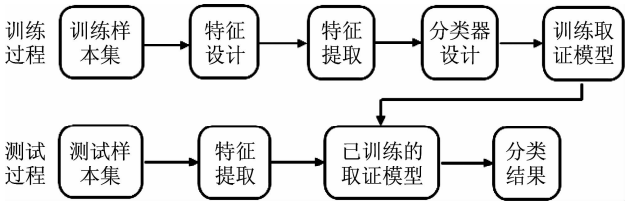

图 2 基于模型的图像取证方法框架图

Fig. 2 The framework of model-based method of image forensics

3 相机源识别方法

相机源识别的主要目的是从数字图像本身获取关于拍摄相机的特征信息(包括相机的品牌、型号或个体)。通过相机源识别可以指出非法图像(如犯罪现场、恐怖主义行为现场等)的所有者,是确保这些数字数据的安全和可信的重要步骤。相机源识别技术通过从数字图像中提取与成像设备相关的特征作为线索加以分析。相机源识别的方法大致包括:传统的基于模型的相机源识别方法和基于深度学习的相机源识别方法,以下将对这两大类方法分别进行介绍。

3.1 传统的相机源识别方法

不同品牌/型号/个体的相机在拍摄数字图像的过程中,由于设备不同带来的对图像处理过程的不同,即使是拍摄同一场景,其生成的数字图像特性也不同。通过提取并分析这些差异特性,可以实现对数字图像相机源的辨识。数码相机的成像过程如图 3 所示。首先,拍摄场景的光线通过相机镜头和彩色滤波器阵列(Color Filter Array, CFA)进入传感器。CFA 使得每个像素点上只保留了一个颜色通道(红色、绿色或蓝色)上的值,剩下的两个颜色通道需要用典型的颜色插值算法进行填充(或拼接)。然后对采集到的信号进行白平衡调整、伽马校

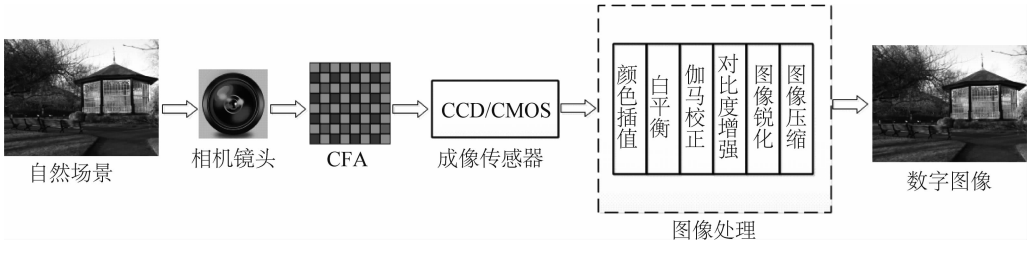

图 3 数码相机成像过程

Fig. 3 The process of capturing image by digital camera

正、对比度增强、图像锐化，有效增强了图像的可视性，生成原始格式的数字图像（例如 TIFF 图像）。一般情况下，数字图像会经过压缩操作（如 JPEG 压缩）保存为压缩格式。由于图像采集过程中硬件设备和图像处理算法的不同，会留下一些只与拍摄相机相关的特征痕迹。通过对这些特征的提取和分类，可以建立图像与拍摄相机之间的联系，实现相机源识别。研究者主要从相机成像的各个环节中入手，挖掘与相机源识别相关的特征，包括与相机透镜失真、相机颜色插值、白平衡、JPEG 压缩、传感器模式噪声相关的特征。

文献[6]提出每台相机都会出现径向畸变现象，通过直线提取方法计算镜头的径向失真系数，建立一个 36 维的向量作为特征，并结合 SVM 分类器来区分三台不同相机拍摄的图像。不同的数码相机一般采用不同的颜色插值算法，因此可以通过估计不同相机的插值系数和插值模式实现相机源识别。文献[7]中使用二次像素相关模型对由颜色插值引起的像素间的相关关系进行建模，对每个颜色通道计算系数矩阵，提取其主成分得到 15 维的特征向量，将其送到前馈反向传播网络中进行相机源识别。测试数据集中的图像由四台相机拍摄而成。文献[8]采用最大期望（Expectation Maximization，EM）算法估计颜色插值系数，将得到的插值系数特征结合 SVM 分类器进行相机源识别。对两种品牌的相机拍摄的图像进行识别，获得的平均检测性能超过 95%。文献[9]首次提出基于自动白平衡（Auto White Balance，AWB）的相机源识别方法，通过图像质量指标（Image Quality Metrics，IQM）估计白平衡参数，结合序列后向特征选择（Sequential Backward Feature Selection，SBS）算法得到一个 404 维的特征向量。在公共数据集 Dresden 上进行实验，对 8 种品牌的相机、17 个不同的相机型号、5 个不同的相机个体进行源辨识。文献[10]提出不同相机在对图像进行 JPEG 压缩时，所采用的 JPEG 量化表可能不同，通过提取 JPEG 量化表可以进行相机源识别；文献[11]探究了将 JPEG 图像的报头信息（包括量化表、霍夫曼码和缩略图等）用以区分不同的相机品牌和个体的可能。以上方法的性能比较如表 1 所示。由于相机成像传感器在制作工艺、材料等方面具有细微的差异，使得每台相机传感器都有其固有的模式噪声。光敏材料的光子响应非均匀性（Photo Response Non Uniformity，PRNU）是由于半导体晶片的非均匀性和缺陷引起的模式噪声，由于其不易消除、在相机寿命周期内相对稳定，可以作为相机源识别的依据。文献[12]完整概述了基于 PRNU 的相机源识别方法，本文中不再赘述。

3.2　基于深度学习的相机源识别方法

基于深度学习的相机源识别方法集中于研究将卷积神经网络（Convolutional Neural Network，CNN）应用于相机源的识别问题。初期的研究主要是对不同的 CNN 结构进行探索，将一些现有的计算机视觉领域的 CNN 结构迁移到相机源识别任务。考虑到相机源识别与计算机视觉任务的不同，越来越多的研究提出在 CNN 的框架中采用预处理的方法，对相机源识别相关特征进行增强。另外，有一些研究关注于其他性能提升的策略，例如通过融合集

表 1　传统的相机源识别方法性能比较

Tab. 1　The performance comparison of traditional method of camera source identification

研究工作	特征模型	数据集	相机源识别任务	性能：Acc
文献[6]	相机镜头的径向失真系数	300 张来自于三台不同相机拍摄的图像	三个不同品牌相机的识别	91.53%
文献[7]	二次像素相关模型	200 张来自于四台不同相机拍摄的图像	四个不同品牌相机的识别	100%
文献[8]	颜色插值系数	280 张来自于两台不同相机拍摄的图像	两个不同品牌相机的识别	95.71%
文献[9]	白平衡参数	Dresden 公共数据集[93]	8 个品牌的相机识别 17 种相机型号识别 5 个相机个体识别	99.26% 98.61% 98.57%
文献[10]	JPEG 量化表	5000 张来自于 10 个相机品牌的 27 个不同的型号相机拍摄的图像	27 种相机型号的识别	>92%
文献[11]	JPEG 图像的报头信息	130 万张来自于 773 台不同的相机和手机拍摄的图像	33 个相机品牌识别 773 个相机个体识别	99% 62%

成多种模型和特征、选择适当的图像块等方法进一步提高深度学习框架在相机源识别上的性能。以下将从网络结构的探索研究、预处理增强特征的方法以及其他性能提升的方法三个方面对基于深度学习的相机源识别研究方法进行介绍。各方法的性能比较如表 2 所示。

(1)网络结构的探索研究

文献[13]提出一个采用 CNN 提取特征、SVM 进行分类的相机型号识别方法。该方法在 64×64 的小图像块上的检测准确率可达 93%。文献[14]进一步探索了利用卷积神经网络进行相机型号识别的可能性。文中研究了不同 CNN 结构、不同训练数据量、不同训练、验证和测试方案对相机源识别准确率的影响。文献[15]提出了一个简单的 CNN 结构,包含三个卷积层以及对应的最大池化层、两个全连接层和 Softmax 分类层,原始的图像被分解为小尺寸图像块进行训练,采用图像块多数投票的策略对整幅图像进行相机源识别。实验结果表明,该方法对 10 种不同相机型号的源辨识准确率可达 99.8%。文献[16]设计了一个具有 13 个卷积层和 3 个全连接层的多分类卷积神经网络。该方法对 JPEG 压缩和加噪等后处理操作具有鲁棒性,但对重采样操作不鲁棒。

表 2 基于深度学习的相机源识别方法性能比较

Tab. 2 The performance comparison of deep learning based method of camera source identification

研究工作	模型	输入尺寸	数据集	相机源识别任务	性能:Acc
文献[13]	CNN+SVM	64 × 64 × 3	Dresden[93]	18 种相机型号的识别	93%
文献[14]	CNN	64 × 64 × 3	Dresden[93]	18 种相机型号的识别	94.93%
文献[15]	CNN+SVM	36 × 36 × 3	Dresden[93]	10 种相机型号的识别	99.9%
文献[16]	CNN	64 × 64 × 3	Dresden[93]	25 种相机型号的识别	93%
文献[21]	ResNet	256 × 256× 3	Dresden[93]	13 个相机品牌的识别 27 种相机型号的识别	99.12% 94.73%
文献[22]	改进的 ResNet	48 × 48 × 3	Dresden[93]	14 个相机品牌的识别 27 种相机型号的识别 74 个相机个体的识别	99.6% 97.1% 52.4%
文献[23]	DenseNet-201	256 × 256 × 1	Dresden[93]	10 种相机型号的识别	98.37%
文献[24]	Inception-Xception	299 × 299	SPC2018[95]	10 种相机型号的识别	93.29%
文献[25]	文献[13] DenseNet-40 DenseNet-121 XceptionNet	64 × 64 × 3 32 × 32 × 3 224 ×224 × 3 299 × 299 × 3	Vision[94]	35 个相机个体的识别	97.47% 95.06% 99.1% 99.31%
文献[26]	有预处理层的 CNN	256 × 256	Dresden[93]	12 种相机型号的识别 14 种相机型号的识别	98.99% 98.01%
文献[28]	有预处理层的 CNN	256 × 256 × 2	Dresden[93]	26 种相机型号的识别	98.58%
文献[29]	有预处理层的 CNN	256 × 256 × 3	Dresden[93]	12 种相机型号的识别 14 种相机型号的识别	98.78% 97.41%
文献[30]	残差卷积神经网络	64 × 64 × 3	Dresden[93] SPC2018[95]	16 种相机型号的识别 12 种相机型号的识别	100% 95.11%
文献[31]	内容自适应的融合残差网络	64 × 64 × 3	Dresden[93]	9 个相机个体的识别	97.03%
文献[32]	多尺度内容无关特征融合网络	64 × 64 × 3	Dresden[93]	23 种相机型号的识别	97.14%
文献[33]	CNN	64 × 64 × 3	Dresden[93]	18 种相机型号的识别	>95%
文献[34]	改进的 VGG	64 × 64 × 3	Dresden[93]	6 个相机品牌的识别 18 种相机型号的识别 74 个相机个体的识别	98.14% 92.62% 41.54%
文献[35]	CNN+ET 分类器	256 × 256 × 3	Dresden[93]	10 种已知相机型号和 15 种未知相机型号的识别	99.38%
文献[36]	孪生网络	256 × 256 × 3	Dresden[93]+自建数据集	65 种相机型号的识别	91.1%

随着网络结构的加深，一些研究工作提出将计算机视觉中采用的先进的 CNN 结构用于相机源识别，如 ResNet[17]、DenseNet[18]、XceptionNet[19]、InceptionNet[20]。文献[21]首次提出将 ResNet 用于相机源识别中，在图像取证的几种不同任务中进行了评估，包括对相机的品牌和型号进行分类。文中还对比了其他计算机视觉领域的 CNN 结构，包括 AlexNet、GoogleNet 和 ResNet 的性能，发现 ResNet 在相机源识别任务中具有更优的性能。文献[22]通过将 ResNet 结构与多任务学习策略相结合，对该方法进行了扩展，进一步提高了性能。将相机品牌、型号和个体的分类等三个取证任务集成到一个 CNN 框架中完成。在 Dresden 数据集上进行实验，对 14 个相机品牌、27 个相机型号和 74 台相机个体进行识别，得到的平均性能分别为 99.6%、97.1% 和 52.4%。文献[23]提出将 DenseNet-201 用于相机型号识别。在 Dresden 数据集上测试了对 10 个相机型号的源辨识性能，平均识别精度超过 98%。文献[24]中提出将预训练的两个 CNN 结构提取的特征作为后续分类 CNN 的输入，其中一个网络采用 InceptionNet 与 ResNet 相结合的结构，另一个网络采用 XceptionNet 结构。考虑三种不同的识别场景，该方法比最好的基准方法在准确率上分别提高0.6%，2.55% 和 1.30%。文献[25]中对比了浅层 CNN 结构[13]、DenseNet 和 XceptionNet 在相机型号识别上的性能，其中 XceptionNet 获得了最佳性能。

(2)预处理增强特征的方法

计算机视觉任务与相机源识别任务不同，计算机视觉任务是基于图像内容本身(强信号)的分类，而相机源识别任务是基于设备属性相关的微弱痕迹进行识别的，而图像内容在相机源识别中被视为一种干扰噪声。因此，需要根据引入预处理技术对取证相关的特征进行增强。

文献[26]针对相机型号的识别任务，提出将 CNN 结构中的第一层设置为一个预处理层用以提取图像残差。在预处理层的设计中，对比了使用高通滤波器(High Pass Filter, HPF)和小波噪声滤波器的性能。实验结果表明，采用高通滤波器作为预处理可以获得更好的性能(准确率提升约 4%)。文献[22]提出采用多尺度的高通滤波器提取三个不同尺度的 HPF 特征，将这些特征合并作为 CNN 的输入并进行最终的相机源分类。受隐写分析特征 SPAM[27]的启发，文献[28]提出了一个有约束的卷积层作为预处理层。在滤波器权值学习更新的过程中设置了这样的约束条件：卷积核的中心权值设置固定为-1，其他位置的权值相加之和为 1。在给定的约束下可以自适应地学习高通滤波器权值。该网络结构用于相机型号的识别，提高了在重采样和重压缩场景下的鲁棒性。除了对预处理层滤波器进行设计的方法外，文献[29]提出首先将图像通过 LBP 编码操作编码为 LBP 映射，再将其输入到 CNN 中进行分类，文中采用一个类似于 AlexNet 的浅层 CNN 结构。对 Dresden 数据集中 12 种相机型号和 14 种相机型号的识别准确率分别达到了 98.78% 和 97.41%。文献[30]中设计了自动的残差提取模块作为预处理模块，用于抑制无关的图像内容信息、提高相机型号识别的性能。

(3)其他性能提升的方法

一些研究通过融合和集成多个模型和特征来提高深度学习方法的源辨识性能。文献[28]在 CNN 的第一层中将有约束的卷积层和中值滤波残差(Median filter residual, MFR)相结合，比单独用有约束的卷积层，获得了更好的性能。文献[31]提出了内容自适应的融合残差网络，实现对小尺寸图像的相机源识别。首先，根据图像内容将图像划分为饱和图像、平滑图像及其他图像三个子集；然后构造不同的残差网络分别对三个图像子集提取有效特征；最后将三个残差网络最后一个残差模块的特征进行融合用于分类。文献[23]利用 DenseNet 提取不同尺寸图像块(64×64、128×128、256×256)的特征，将其融合后输入分类模块得到最终的识别结果。文献[24]采用 InceptionNet 与 ResNet 相结合的结构使得源辨识的性能得到了提高。文献[32]提出了一个多尺度内容无关特征融合网络(Multiscale content-independent feature fusion network，MCIFFN)。该网络由三个并行分支网络组成，在其中两个分支网络中设计了自适应的滤波模块用于滤除图像内容、提取噪声特征。第三个分支是直接对图像进行特征学习。最终将三个分支网络中提取到的不同尺度的内容无关特征进行融合用于相机源识别。

在平滑的、高亮度的非饱和区域更有利于学习和提取相机源识别取证特征，因此，采用图像块选择的方法可以提高取证性能。一些研究正是基于这样的分析，提出了图像块的选择方法，通过选取

更有利于 CNN 训练的图像块提升其源辨识取证的性能。文献[13]中只选择了均值接近图像动态范围一半的图像块进行训练。文献[14]给出了另一种选择图像块的准则,即在保证均值接近图像动态范围一半的情况下找到一些更能显示统计差异的纹理图像块。在这种方法中,像素块的质量值由其方差和平均值计算。选择高质量值的像素块来训练 CNN 模型。文献[33]提出对每个像素块估计一个代表相机模型属性可靠性的值,作为衡量标准,用来挑选用于训练 CNN 的图像块。文献[34]提出根据局部纹理和语义标准选择具有代表性的图像块,可以使深度神经网络更好地学习相机源识别相关的特征,提高网络的鲁棒性和泛化能力。

另一些深度学习方法是针对更具挑战性的场景提出的,通过改进分类器实现相机源识别性能的提升。上述研究工作都是在解决“闭集”上的图像相机源识别问题。所谓“闭集”指的是假设测试数据中涉及的相机类型在训练数据集中出现过,意味着用于采集训练图像的相机集合完全包含了采集测试图像的相机集合。然而,现实应用场景中更常见的是“开集”上的检测,即待测图像可能是来源于未知类型的相机。针对这一场景,文献[35]提出了两种不同的方案来解决“开集”的相机型号识别问题,其目的是判断采集待测图像的相机是已知的还是未知的。第一种方法使用置信值映射来代替分类层,并使用阈值策略来评估该相机型号是已知还是未知。另一种方法是在 CNN 提取特征后使用不同的分类器。文献[36]提出采用孪生网络来衡量图像之间的相机源相似性。从 CNN 的最后一层提取特征并输入孪生网络学习相机源相似性的度量,从而验证两幅待测图像是否由同一相机采集。

4 计算机图形学方法生成图像取证

计算机图形学方法生成(CG)图像取证的目的是判断一幅待测图像是自然图像还是利用计算机图形学方法生成的图像。随着计算机图形学技术的发展,通过计算机软件(如 3Ds MAX、Maya、Softimage-xsi 等)可以生成现实中不存在的场景,制作出的生成图像在视觉效果上可以与自然图像相差无几,达到以假乱真的效果。计算机图形学方法生成图像的过程如图 4 所示。首先利用计算机软件构建一个 3D 多边模型模拟期望生成图形的形状;将 3D 模型投影为二维图像的形式表示,通过消隐技术消除被遮挡的不可见的线或面;然后采用纹理生成技术为模型表面赋予几何纹理或颜色纹理;最后建立适当的光照模型,模拟真实环境下的光源照射物体时达到的光照效果(如漫反射、镜面反射及透射等)。自然图像和 CG 图像在生成的原理上有很大不同。自然图像是通过图像采集设备将真实场景进行投影而形成的照片;而 CG 图像的产生是通过计算机软件和图像处理方法对照片形成过程的模拟,因此其在统计特性(如直方图的连续性、细小纹理的复杂程度等)上会与自然图像存在差异。CG 图像取证技术正是利用这些差异来区分这两种图像。

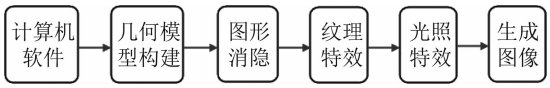

图 4 计算机图形学方法生成图像的过程

Fig. 4 The generating process of CG images

传统的 CG 图像取证方法是在空间域或变换域中通过手工设计特征来检测的,不同特征的性能比较如表 3 所示。文献[37]中根据自然图像与 CG 图像的物理生成过程的差异,提出了一种几何图像模型,在最细尺度上采用分形几何,在中等尺度上采用微分几何。基于几何的方法可以挖掘自然图像与 CG 图像具有的独特物理特征,如伽马校正在自然图像中留下的痕迹以及 CG 图像成像过程由于 3D 建模引入的特征。文献[38]提出将图像特征函数和小波子带的统计矩作为特征,并比较了在 RGB 和 HSV 两种不同颜色空间中提取的特征,结果表明采用 HSV 颜色空间中构造的特征在 CG 图像取证上表现出更好的性能。文献[39]通过检测原始自然图像的颜色插值痕迹,以区分自然图像和 CG 图像,取得了良好的取证性能。但由于一些后处理操作(如尺寸调整)可能会消除颜色插值结构,因此该方法对这些后处理操作敏感。文献[40]提出利用局部二值模式(Local binary patterns, LBP)统计直方图特征来识别自然图像与 CG 图像,取得了良好的取证性能。然而该方法性能也容易受到图像尺寸调整操作的影响。文献[41]提出了一阶小波统计量和高阶小波统计量相结合的特征,用于自然图像和非真实的 CG 图像的识别。文献[42]在前人研究的基础上,提出了一组组合特征,包括基于颜色直方

表 3　传统的计算机图形学方法生成图像的取证方法性能比较

Tab. 3　The performance comparison of traditional method of CG image forensics

研究工作	特征模型	数据集	性能
文献[37]	物理特征	Columbia[96]	Acc=83.5%
文献[38]	图像特征函数和小波子带的统计矩	Columbia[96]	Acc=82.1%
文献[39]	自然图像的颜色插值痕迹	Columbia[96]	Acc=98.4%
文献[40]	LBP 统计直方图特征	2455 张自然图像,2455 张 CG 图像	Acc=98.3%
文献[41]	一阶小波统计量和高阶小波统计量相结合的特征	40000 张自然图像,6000 张 CG 图像	FPR=0.8%
文献[42]	组合特征	Columbia[96]	Acc=90%
文献[44]	图像局部边缘块统计特性	1000 张自然图像,900 张 CG 图像	Acc=95.7%
文献[45]	残差图像的直方图统计量和多重分形谱与回归模型的适应度	3000 张自然图像,3000 张 CG 图像	Acc=98.69%
文献[46]	隐写分析特征	15200 张自然图像,7492 张 CG 图像	Acc=87.6%

图特征[43]、HSV 颜色空间中基于矩的统计特征[38]、局部块统计特征[37]、基于纹理插值的特征。文献[44]提出了一种分析图像局部边缘块统计特性的方法。首先利用 Voronoi 单元确定关键采样点,构造图像局部边缘的视觉词汇;然后利用视觉词汇的直方图形成特征向量,训练 SVM 分类器进行分类。文献[45]利用线性回归模型提取高斯低通滤波残差图像,并将残差图像的直方图统计量和多重分形谱与回归模型的适应度相结合作为特征来区分自然图像和 CG 图像。另外,还有一些检测 CG 图像的方法[46]是借鉴隐写分析特征提出的。

目前,已有越来越多的研究提出了基于深度学习的 CG 图像取证方法。文献[47]评估了 VGG 结构用于 CG 图像取证的性能,实验发现去除 VGG 结构中的 Max-pooling 层可以提高检测性能。因此,文中基于 VGG 结构进行改进,提出了一种没有任何池化层的六层 CNN 结构,在 32×32 的图像块上实现了超过 98% 的检测准确率。文献[48]在 VGG-19 和 ResNet-50 结构上测试了几种训练策略。结果表明,在训练阶段使用迁移学习并且采用微调的 ResNet-50 模型可以获得最好的性能。在 DSTok 数据集上,平均检测准确率约为 96.1%。文献[49]提出了一个双通道的 CNN 和 RNN 混合的框架用于检测 CG 图像。首先进行颜色空间转换,将图像从 RGB 颜色空间转换为 YCbCr 颜色空间;然后利用 Schmidt 滤波器组对亮度分量进行处理,得到 13 种不同的滤波响应;最后分别将色度分量 Cb、Cr 和亮度分量滤波后的响应送入 CNN 和 RNN 的混合框架。提出的方法与文献[47]中的方法相比,检测精度提高了 4%。文献[50]中设计了一个五层 CNN,使用隐写分析领域的高通滤波器对输入图像进行预处理。使用了三种高通滤波器:SQUARE5x5、SQUARE3x3 和 EDGE3x3。在收集的 1800 张 CG 图像和 1800 张自然图像上进行实验,检测准确率达到 100%。文献[51]设计了一个没有预处理层的 CNN 网络,在训练阶段可以自适应地学习卷积滤波器的权值。采用“局部-全局”策略,先对小块图像进行训练再通过简单的多数投票对整张图像进行分类,该方法的性能优于文献[45]等传统的方法,表明了基于 CNN 的方法在 CG 图像取证上的有效性。文献[52]提出了一种新的统计特征提取(Statistical features extraction, SFE)层,并将其嵌入最后一层卷积层和第一个全连接层之间。SFE 层提取四种统计特征,包括均值、方差、最大值和最小值。文中还对比了端到端的 CNN 的性能和采用不同的分类器(如 LDA 和 SVM)的性能,结果表明采用端到端的 CNN 模型得到了更好的结果。文献[53]采用 VGG-19 作为特征提取器,提取前三个最大池化层之后的卷积层输出并计算其均值和方差,得到最终特征,然后将这三组最终特征输入分类器得到检测结果。提出的方法优于对比的方法,在高分辨率的图像上检测率可达到 100%。文献[54]探究了使用不同网络结构作为特征提取搭配不同的分类器(Softmax、k 最近邻、XGBoost 和 SVM)的性能。最后选择 ResNet-50 作为特征提取,与具有 RBF 核的支持向量机分类器相结合,获得最好的性能。文献[55]中设计了一个具有自编码模块的 CNN 用于 CG 图像取证。该模块以彩色图像为输入,提取颜色通道之间的相关性。实

验结果表明该网络在分类性能方面优于对比的 CG 图像取证方法。文献[56]提出了一种基于注意的双分支卷积神经网络(Attention-based dual-branch convolutional neural network,AD-CNN)。该网络采用并行的双分支结构,每个分支具有相同的 CNN 网络结构,但第一个卷积层具有不同的卷积核大小,从而提取多尺度下的浅层次取证痕迹。通过基于注意力的融合模块对各分支的输出特征进行联合优化,自动为不同分支分配非对称权值。该方法在小尺寸图像块的检测上具有明显的优势。以上基于深度学习的 CG 图像取证方法的性能比较如表 4 所示。

5 AI 合成图像取证

近年来,随着人工智能(Artificial Intelligence, AI)技术的快速发展和广泛应用,一些基于 AI 技术的图像合成方法已经被开发出来,用以生成虚假的多媒体内容信息。这些方法主要包括自编码器(Autoencoder)和生成对抗网络(GAN),统称为“深度伪造”(Deepfake)。深度伪造技术可以生成的图像内容比计算机图形学方法生成的图像内容要真实得多。这给取证技术带来了极大的挑战。大量的研究和竞赛关注于深度伪造多媒体的检测[57],例如 NIST 和 Facebook 分别推出了 MFC2018 和 DFDC 两个用于深度伪造检测的数据库,并发表了一些关于这方面的研究调查。文献[58]概述了媒体取证和深度伪造的主要方法。

基于 AI 技术的图像合成方法中,采用生成对抗网络(Generative Adversarial Network, GAN)合成图像是最具代表性的方法。GAN 是一个基于博弈论的生成模型学习框架,由生成器 G 和判别器 D 两部分组成,其基本框架如图 5 所示。生成器 G 的输入为随机噪声 z,输出为生成样本 $G(x)$;判别器 D 的输入为生成样本 $G(x)$和真实样本 x,输出是一个概率值,概率越大表明生成样本 $G(x)$越接近真实分布。输出概率值反馈给生成器 G,用于指导 G 的训练。训练过程中,生成器希望生成样本 $G(x)$接近真实样本,即希望概率尽可能大;而判别器的目标是尽可能区分真实样本和生成样本。当通过零和博弈达到纳什均衡时,GAN 模型达到最优。在基本的 GAN 框架上衍生出了很多先进的 GAN 结构。由于高分辨率图像容易导致判别器产生梯度问题,早期 GAN 模型仍难以生成高分辨率图像。对此,文献[59]提出了 PGGAN,首次生成了 1024×1024 高分辨率图像。在 PGGAN 的基础上,文献[60]提出了基于风格的 GAN 模型(即 StyleGAN)。该模型可以生成比 PGGAN 更加逼真的高分辨率图像。GAN 模型经过不断发展已经能够生成逼真的高分辨率图像,但生成图像内容具有较大随机性,难以控制其语义类别。文献[61]提出的 CycleGAN 具备在内容生成过程中考虑类别信息的特点,成功用于非配对数据的图像风格迁移。然而 CycleGAN[61] 只能解决单风格迁移问题。文献[62]进一步提出了 StarGAN 来解决多风格图像相互转换的问题。文献[63]将自注意力模块引入 GAN 模型,提出了一种基于自注意力机制的 GAN 模型(Self-Attention Generative Adversarial Network, SAGAN)。自注意力机制能够更好的捕捉

表 4 基于深度学习的计算机图形学方法生成图像的取证方法性能比较

Tab. 4 The performance comparison of deep learning based method of CG image forensics

研究工作	模型	输入尺寸	数据集	性能:Acc
文献[47]	改进的 VGG	32 × 32× 3	Columbia[96]	98.2%
文献[48]	微调的 ResNet-50	224 × 224	DSTok[97]	96.1%
文献[49]	CNN 和 RNN 混合框架	96 × 96	3Dlink[49]	94.87%
文献[50]	有预处理层的 CNN	650 × 650	WIFS[52]	100%
文献[51]	CNN	233 × 233	Columbia[96]	93.20%
文献[52]	具有 SFE 层的 CNN	100 × 100 × 1	WIFS[52]	93.2%
文献[53]	VGG-19+LDA 和 SVM 分类器	–	WIFS[52]	99.89%
文献[54]	ResNet-50+SVM 分类器	224 × 224 × 3	DSTok[97]	94%
文献[55]	具有自编码模块的 CNN	96×96	3Dlink[49]	94.18%
文献[56]	AD-CNN	32 × 32 × 3 64 × 64 × 3	3Dlink[49]	83.75% 87.82%

全局信息，并且保持良好的运算效率。在 SAGAN 的基础上，文献[64]引入正交正则化的思想，提出了一种新的 GAN 模型，即 BigGAN。BigGAN 增加了批大小和网络宽度（即特征图通道数量），从而显著提升生成图像的真实性评价指标。随着基于 GAN 的图像合成技术的进步，有越来越多的研究关注于 GAN 合成图像的取证。这些取证方法大致可以分为：基于手工设计特征的方法和基于深度学习的方法。

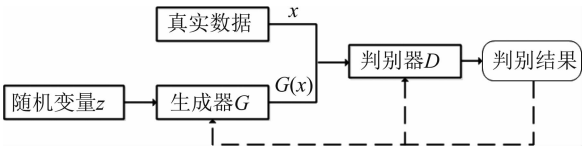

图 5　生成对抗网络的基本框架

Fig. 5　The framework of generative adversarial network

手工设计的特征主要包括图像曝光特性、颜色特征、噪声特征等，不同方法的性能比较如表 5 所示。文献[65]分析指出相机拍摄的图像通常会具有饱和或曝光不足的区域，而 GAN 在合成图像的过程中，由于生成器中有归一化的操作，因此 GAN 合成的图像缺少这些区域。文中提出一种红绿双变量直方图和一种异常曝光像素比例的取证特征，结合 SVM 分类器进行检测。文献[66]针对 GAN 合成图像过程中由于上采样操作引入的生成痕迹，通过理论分析说明了这种痕迹表现为频域的棋盘效应，由此提出了一个基于频谱输入而不是像素输入的分类器模型。该方法具有较强的通用性，对于在训练过程中没有出现的 GAN 模型合成的图像，也能取得很好的检测性能。文献[67]通过对不同颜色空间中相邻像素间相关性的评估，分析发现自然图像和 GAN 合成图像在颜色空间中存在差异，从而提出了一种基于颜色分量统计特性的特征集。首先，将彩色图像转换为 RGB、HSV 和 YCbCr 不同的颜色空间。然后分别生成 R、G、B、H、S、V、Y、Cb、Cr 通道的残差图像，并计算所有残差图像的共生矩阵。实验评估了不同检测情况下的性能，结果表明当训练和测试数据不匹配时，该方法也能够准确识别 GAN 合成图像，性能优于现有方法；此外当 GAN 模型未知、采用真实图像进行训练时，该方法也取得了良好的性能。

在基于深度学习的方法中，文献[68]评估了四种图像取证方法（包括 CG 图像取证方法[52]、隐写分析方法[69]、篡改检测方法[70]和多目标图像处理操作检测方法[71]）和计算机视觉领域里常用的几种 CNN 结构对 GAN 合成图像检测的性能，对比的 CNN 结构包括：DenseNet[18]，InceptionNet v3[19]，XceptionNet[20]和 Cycle-GAN[61]中的判别器。实验结果表明，XceptionNet 获得了最高的平均检测精度，对图像压缩操作（如 Twitter 压缩）鲁棒。文献[72]中考虑了两种检测 GAN 合成图像的方法。第一种方法是在 GAN 模型已知的情形下，使用 GAN 中的鉴别器来检测生成的图像。第二种方法是 GAN 模型未知的情形下，评估了基于人脸质量评价、Inception 评分和 VGG 特征的性能。实验结果表明，两种方法都能够有效检测 GAN 合成的图像。在第二种方法中，基于 VGG 特征的检测器获得了最好的性能。文献[73]提出残差域中可以反映出自然图像和 GAN 合成图像之间的主要差异。因此，将 CNN 第一层设置为预处理层，采用高通滤波预处理生成残差特征。该方法能有效地识别 PG-GAN[59]合成的高质量假脸图像。文献[74]提出一种结合共生矩阵和深度学习的方法来检测 GAN 合成的图像。使用输入图像的三个通道分别计算共生矩阵，然后将其输入 CNN 中进行分类。该方法对于 CycleGAN[61]和 StarGAN[62]合成图像检测的分类准确率达到 99% 以上。文献[75]中尝试构建一个通用检测器，检测由 GAN/CNN 生成的图像，而不考虑具体的结构和使用的数据

表 5　传统的 AI 合成图像取证方法性能比较

Tab. 5　The performance comparison of traditional method of AI synthesized image forensics

研究工作	特征模型	AI 合成方法	数据集	性能
文献[65]	红绿双变量直方图和异常曝光像素比例的取证特征	PGGAN	MFS2018 子数据集[98]	AUC = 0.7
文献[66]	频域的棋盘效应特征	CycleGAN	自建数据集	Acc = 97.2%
文献[67]	颜色分量差异特征	DCGAN、WGAN-GP、PGGAN、StyleGAN 等	自建混合数据集	Acc = 99.74%

库。文中收集了11种基于GAN/CNN生成模型(包括PG-GAN[59], StyleGAN[60], CycleGAN[61], StarGAN[62]等)合成的图像数据集。网络结构采用Resnet-50,经过设计适当的预处理、后处理和数据增强,可以对多种生成模型、训练中没有用到的数据集都具有很好的通用性能。文献[76]采用两阶段学习的方法。首先采用三元损失(Triplet loss)从不同GAN模型合成的图像中学习通用伪造特征(Common fake feature);然后用具有不同卷积核尺寸的残差模块提取局部和全局的通用伪造特征,对分类层进行微调。该方法可以有效用于GAN合成假人脸图像的检测。文献[77]利用GAN生成器设计中存在的缺陷,即采用反卷积操作进行上采样的过程中会造成生成图像中全局信息的丢失,设计了一个具有自注意力(Self-Attention)机制的CNN检测器。该方法可以准确地捕捉由该缺陷造成的特殊纹理模式,与现有的CNN方法相比,在检测精度有显著提高。文献[78]通过分析GAN生成图像的频率域(例如:离散余弦变换,Discrete Cosine Transform,简称DCT),发现在中高频分量存在明显的尖峰。文中设计了一种基于DCT频域变换以及卷积神经网络的检测方法。采用了浅层卷积神经网络和残差神经网络分别作为分类器。实验结果表明,将经过预处理的频谱输入卷积神经网络,相比于直接将图像像素作为输入,能取得更好的检测结果。文献[79]提出一种基于颜色通道频谱及胶囊网络的GAN图像检测算法。该算法首先计算输入图像R,G和B三个颜色通道的离散傅里叶变换(Discrete Fourier Transform,DFT)获得对应频谱。然后将三个颜色通道的频谱按通道维度进行堆叠,输入胶囊网络中获得最终检测结果。各方法的性能比较如表6所示。

表6 基于深度学习的AI合成图像取证的取证方法性能比较

Tab. 6 The performance comparison of deep learning based method of AI synthesized image forensics

研究工作	AI合成方法	数据集	模型	性能:Acc
文献[68]	CycleGAN	自建数据集	CycleGAN判别器	83.58%
			隐写分析方法	94.40%
			篡改检测方法	95.07%
			多目标图像处理操作检测方法	84.86%
			CG图像取证方法	85.71%
			DenseNet	89.19%
			InceptionNet v3	89.09%
			XceptionNet	94.49%
文献[72]	DCGAN、WGAN	自建数据集	DCGAN判别器	95.51%
			VGG+FLD	>90%(DCGAN) >94%(WGAN)
文献[73]	PGGAN	自建数据集	Lap-CNN	96.3%
文献[74]	CycleGAN、StarGAN	自建数据集	结合共生矩阵的CNN	>99%
文献[75]	11种基于GAN/CNN生成模型	自建数据集	改进的Resnet-50	93.0%
文献[76]	DCGAN、WGAP、WGAN-GP、LSGAN、PGGAN	自建数据集	耦合深度神经网络	98.6%(DCGAN)、 89.5%(WGAP)、 87.6%(WGAN-GP)、 98.1%(LSGAN)、 95.1%(PGGAN)
文献[77]	PGGAN	自建混合数据集	具有自注意力机制的CNN	99.3%
文献[78]	StyleGAN	自建数据集	DCT频域变换+CNN	100%
文献[79]	StyleGAN	自建数据集	胶囊网络	>99%

6　重获取图像的取证

重获取图像的产生过程如图 6 所示。原始的自然图像首先经过一次中间媒介的映射，中间媒介包括手机屏幕、PC 机 LCD 显示器、投影仪、打印机等；然后再次利用图像采集设备（包括手机、相机、扫描仪等）对其重新获取得到的图像。重获取过程常常被用作消除或减弱图像篡改痕迹的手段，因此对图像的重获取检测可以成为图像篡改取证的辅助手段；另外，重获取图像取证可以用于判别原始照片的所有者，具有版权保护的功能；重获取图像取证还可以有效防止在人脸识别中使用肖像照片进行认证的欺诈行为，提高生物识别技术的稳定性。

对于不同中间媒介获取的图像，研究者根据原始自然图像和重获取图像的统计特性差异，提出了各种算法。基于手工设计特征的传统方法的性能比较如表 7 所示。针对重获取打印图像的检测问题，文献[80]计算了图像镜面反射分量在图像强度中的占比，发现自然图像的占比梯度直方图分布近似服从 Laplacian 分布，而重获取图像的占比梯度直方图分布呈现 Rayleigh 分布。文献[81]提出了一种基于物理的单视图重新捕获图像检测方法，基于物理

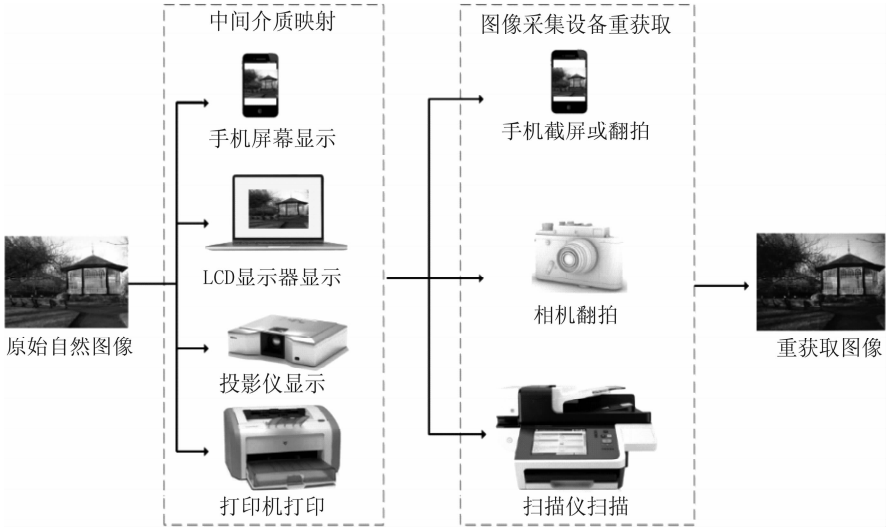

图 6　重获取图像的产生过程

Fig. 6　The generating process of recaptured images

表 7　传统的重获取图像取证方法性能比较

Tab. 7　The performance comparison of traditional method of recaptured image forensics

研究工作	特征模型	数据集	性能：Acc
文献[81]	基于物理的特征集	608 张自然图像，589 张重获取图像（后置摄像头） 420 张自然图像，410 张重获取图像（前置摄像头）	95.0%（后置摄像头） 91.3%（前置摄像头）
文献[82]	基于马尔可夫过程的离散余弦变换（DCT）系数阵列特征	3994 对自然图像和重获取图像	99.33%
文献[83]	LBP、MSWS、CF	2000 张自然图像，2700 张重获取图像	99.5%
文献[84]	噪声特性、JPEG 压缩特性	2400 对自然图像和重获取图像	98.58%
文献[85]	边缘模糊特性	1035 张自然图像，2520 张重获取图像	99%
文献[86]	边缘轮廓的差异	900 张自然图像，1440 张重获取图像	99.88%（自然图像） 99.99%（重获取图像）
文献[87]	JPEG 压缩特性、小波分解特征	636 对自然图像和重获取图像	98.32 %
文献[88]	灰度共生矩阵	400 对自然图像和重获取图像	92.5%（Data A） 97.3%（Data B） 93.3%（Data C）

的特征集由背景信息、与纸张表面相关的镜面反射分量的分布、反映重获取图像渲染过程中独特属性的图像梯度、颜色信息、对比度以及模糊度等。文献[82]中研究发现由于图像细节信息的丢失,使得重获取图像的图像质量有所下降,即出现了视觉上的模糊。针对此特性提出了基于马尔可夫过程的离散余弦变换(DCT)系数阵列特征来表征这种变化。

针对PC机LCD显示器的重获取图像检测,文献[83]提出了三种统计特征,即局部二值模式(LBP)、多尺度小波统计(Multi-scale wavelet statistics, MSWS)和颜色特征(Color features, CF)来检测高质量的重获取图像。文献[84]提出一种基于噪声和双JPEG压缩痕迹的检测方法。在噪声特性方面,首先采用小波阈值方法对图像进行去噪,将原图像与去噪后图像做差值得到噪声特征;在JPEG压缩特性方面,将重获取图像看作是经过双重JPEG压缩,通过分析DCT系数不同模式中第一个有效数字的概率得到特征。文献[85]提出了一种基于学习边缘模糊的LCD显示器重获取图像检测算法。利用k-奇异值分解方法从自然图像和重获取的图像中选择边缘的线扩展轮廓训练两组字典,然后利用字典近似误差和训练图像的平均边缘扩展宽度作为SVM分类器特征。文献[86]也提出图像中的边缘是包含详细信息的显著特征。为了分析原始自然图像和重获取图像的边缘轮廓的差异,文中采用平稳小波变换(Stationary wavelet transform, SWT)方法分别获取水平、垂直和对角线上的边缘,对边缘轮廓的方向和像素个数进行了计算。通过分析边缘轮廓,挖掘原始自然图像和重获取图像的差异。文献[87]针对高分辨率自然图像和高质量LCD显示器重获取图像的鉴别,提出了两种有效的特征,包括:JPEG压缩产生的块效应和模糊效应以及带有混叠增强预处理的小波分解特征。文献[88]提出一种基于灰度共生矩阵的新方法用于检测LCD显示器重获取图像。为了分析自然图像与重获取图像的差异,该方法首先通过子波变换,提取高频和低频的信息,并在此基础上计算了相对灰度共生矩阵,将灰度共生矩阵作为特征,利用SVM对图像进行分类。

目前已有一些基于深度学习的重获取图像取证方法提出,各方法的性能比较如表8所示。文献[89]是第一个尝试使用深度学习检测重获取图像的工作。该研究主要是针对小尺寸的重获取图像取证,提出了Laplacian卷积神经网络(L-CNN)。在该方法中,将Laplacian滤波器嵌入到CNN的第一层,以提取增强的取证相关信息。文中对五种不同的高通滤波器进行了评价。实验结果表明,使用Laplacian滤波器得到的性能优于使用其他高通滤波器或不使用滤波器得到的性能。L-CNN在64×64的小尺寸图像块上能达到96%的检测准确率。文献[90]提出一个九层的CNN结构,在64×64的图像块上进行分类,使用多数投票策略得到对整幅原图像的检测结果。该方法与传统的重获取图像取证方法相比略有改进。文献[91]提出了一种将卷积神经网络与递归神经网络(Recursive neural network, RNN)结合起来的新框架。首先将图像块作为卷积神经网络的输入提取局部块内部的特征,然后再采用递归神经网络来提取相邻块之间的相关性。

表8 基于深度学习的重获取图像取证的取证方法性能比较

Tab. 8 The performance comparison of deep learning based method of recaptured image forensics

研究工作	模型	输入尺寸	数据集	性能:Acc
文献[89]	L-CNN	N × N × 3	NTU-Rose[83]、LCD_R[92]	99.74%(512×512) 99.74%(256×256) 98.48%(128×128) 95.23%(64×64)
文献[90]	CNN	64 × 64 × 3	ICL[85]	96.60%
文献[91]	CNN+RNN	32 × 32 × 3	ASTAR[99] NTU-Rose[83] ICL[85]	93.29%(ASTAR) 98.67%(NTU-Rose) 99.54%(ICL)
文献[92]	CNN	64 × 64 × 1	LS-D[92]	99.90%

与传统的方法和其他基于深度学习的方法相比，该方法可以获得更好的性能。文献[92]提出了一个用于评估重新捕获的图像取证技术的大型数据集。该数据集包括 14500 张重新获取的图像和 14500 张原始图像。这些图像是由各种设备（如照相机，显示器，扫描仪，打印机）采集的。文中提出了一个八层的 CNN 结构，在第一层中设计了 16 种不同的高斯滤波器提取残差。对于 64×64 图像块的检测准确率达到 99.9%，这与传统的方法相比，性能上获得了极大的提升。

7　评价指标和数据集

本节介绍在图像源辨识取证中常用的评价指标和数据库。

在大部分图像源辨识取证研究中都用到的性能衡量标准为准确率（Acc），定义式如下：

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \tag{1}$$

其中 TP 和 TN 分别表示正确分类为正样本和负样本的数目，FP 和 FN 分别表示错误分类为正样本和负样本的数目。准确率代表整体的分类准确程度，既包括正样本，也包括负样本。虽然准确率可以判断总的正确率，但是在样本不平衡的情况下，并不能作为很好的指标来衡量结果。因此也就衍生出了其他两种指标：精准率和召回率

精确率（Precision），又称为查准率。精确率的含义是所有分类为正的样本中实际为正的样本的比率。其代表了对正样本的分类准确程度，定义式如下：

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{2}$$

召回率（Recall），又称为查全率。召回率的含义是实际为正的样本中被分类为正样本的概率。定义式如下：

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{3}$$

为了同时考虑精确率和召回率，让两者取得平衡，通常采用 F1 分数作为指标。F1 分数被定义为精确率和召回率的调和平均数，其表达式为：

$$\mathrm{F1}=2\times\frac{\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{4}$$

图像源辨识取证研究中另外两个重要的性能指标为 ROC（Receiver Operating Characteristic）曲线和 AUC（Area Under Curve）曲线。ROC 曲线横坐标为假正率（FPR），纵坐标为真正率（TPR）。假正率（FPR）表示的是实际为负的样本被分类为正样本的概率。真正率（TPR）表示的是实际上是正的样本被分类为正样本的概率。FPR 和 TPR 的定义式如式（5）、（6）所示。

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \tag{5}$$

$$\mathrm{TPR}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{6}$$

ROC 曲线展示了当改变模型中的阈值时，FPR 和 TPR 的变化关系。AUC 曲线表示的是 ROC 中曲线下的面积，是对所有可能的分类阈值的效果进行综合衡量。AUC 的值一般是介于 0.5 和 1 之间，越接近于 1 表示模型的性能越好。

图像源辨识中常用的开源数据集如表 9 所示。下面分别对相机源识别、CG 图像取证、GAN 合成图像取证以及重获取图像取证中常用的开源数据集进行介绍。

用于相机源识别的公共数据集主要包括 Dresden[93]、Vision[94] 和 SPC2018[95]。Dresden 数据集[93]发布于 2010 年，包含超过 14000 张室内外场景的照片，由来自 25 个不同型号的 73 台相机拍摄。目前，Dresden 是相机源识别中应用最为广泛的数据集。VISION[94] 发布于 2017 年，是用于设备源辨识的视频和图像数据集。它包含 34427 张图像和 1914 段视频组成，既有原始格式的图像和视频，也有社交版本（Facebook、YouTube 和 WhatsApp）的图像和视频，来自 11 个品牌的 35 个设备个体。SPC2018[95] 是以相机型号识别为主题的数据集，2018 年由 IEEE 信号处理杯竞赛发布。该数据集包含了来自 10 种不同相机型号（包括傻瓜相机、手机相机和数码单反相机）的 2750 张图像，每种相机型号分别拍照了 275 张图片。

用于 CG 图像取证的数据库主要包括：Columbia[96]、DSTok[97]、WIFS[52] 以及 3Dlink[49]。Columbia[96] 包含 800 张从 40 个 3D-图形网站（如：www.softimage.com，www.3ddart.org，www.3d-ring.com 等）下载的 CG 图像，1200 张来自个人收藏的自然图像，800 张从谷歌上收集的自然图像，以及 800 张重获

表 9 图像源辨识取证开源数据集

Tab. 9 Open source dataset of image source identification forensics

数据集	用途	描述	大小
Dresden[93]	用于相机源识别	由来自 25 个不同型号的 73 台相机拍摄的室内外场景的照片组成	14000 余张图像
Vision[94]	用于相机源识别	由来自 11 个品牌的 35 个设备个体拍摄的图像和视频组成	34427 张图像，1914 段视频
SPC2018[95]	用于相机源识别	由来自 10 种不同相机型号的图像组成	2750 张图像
Columbia[96]	用于 CG 图像检测	包括来自于个人收藏和谷歌下载的自然图像，来自于 40 个 3D-图形网站的 CG 图像以及重获取的 CG 图像	2000 张自然图像，1600 张 CG 图像
DSTok[97]	用于 CG 图像检测	所有的 CG 图像和自然图像都是从互联网上收集的，都经过了 JPEG 压缩	4850 对自然图像和 CG 图像
WIFS[52]	用于 CG 图像检测	CG 图像来自于关卡设计参考数据集，自然图像来自于 RAISE 数据集	1800 对自然图像和 CG 图像
3Dlink[49]	用于 CG 图像检测	包含 3Dlink 网站上下载的 CG 图像和使用不同型号相机在不同环境条件下拍摄的自然图像	6800 对自然图像和 CG 图像
MFS2018 子数据集[98]	用于 GAN 合成图像	MFC18 GAN Full 和 GAN crop set 为 MFS2018 子数据集，GAN 合成图像主要的操作包括换脸、填充、擦除等	2340 张 GAN 生成图像
LS-D[92]	用于重获取图像检测	包含四种类型的重获取图像：打印图像的照片，打印图像的扫描件，LCD 显示器图像的翻拍，LCD 显示器图像的截屏	145000 对自然图像和重获取的图像
NTU-Rose[83]	用于重获取图像检测	从 LCD 显示器中重获取得到的高质量图像的集合	300 自然图像，2700 张重获取图像
ICL[85]	用于重获取图像检测	由 LCD 显示器中重获取得到的图像组成	1035 张自然图像，2520 张重获取图像
ASTAR[99]	用于重获取图像检测	智能手机拍摄的重获取图像数据集，数据集分为三个子集	子集 A：1094 张自然图像，1137 张重获取图像。 子集 B：1094 张自然图像，1765 重获取图像。 子集 C：587 对自然图像和重获取的图像组成。

取的 CG 图像。DSTok[97] 包含从互联网上收集的 CG 和真实照片（Photograph，PG）图像。有 4850 对 CG 和 PG 图像。它们经过 JPEG 压缩的，文件大小在 12KB 到 1.8 MB 之间。WIFS[52] 是包含有 1800 张 CG 图像和 1800 张 PG 图像。CG 图像是从关卡设计参考数据集（Level Design Reference Database）上下载的。该数据库包含 60000 多张 JPEG 格式的高分辨率电子游戏截图，其中只有 5 种电子游戏的截图比较接近于真实图像，可以被纳入 WIFS，因此只有 1800 张 CG 图像被选中。PG 图像来源于 RAISE 数据集，是相机拍摄的高分辨率真实图像，直接转换为 JPEG 格式。3Dlink[49] 包含 3Dlink 网站上下载的 6800 张 CG 图像和使用不同型号相机在不同环境条件下拍摄的 6800 张 PG 图像。

用于 GAN 合成图像检测的公共数据集有 MFS2018 子数据集[98]。MFS2018[98] 是 2018 年发布的媒体取证挑战数据集，旨在帮助推进图像和视频取证技术的发展。数据库包括超过 176000 张高质量的图像和 11000 个高质量视频，涉及的图像和取证任务多达 7 个。其中用于检测 GAN 合成图像这一取证任务的数据集为 MFC18 GAN Full 和 GAN crop set，数据集分别包含 1340 张图像和 1000 张图像，GAN 合成图像主要的操作包括换脸、填充、擦除等。

用于重获取图像检测的公共数据集主要包括：LS-D[92]、NTU-Rose[83]、ICL[85] 和 ASTAR[99]。LS-D[92] 是一个用于评估重获取图像取证性能的大规模数据库，包含四种类型的重获取图像：（1）打印图像的照片；（2）打印图像的扫描件；（3）LCD 显示器图像的翻拍；（4）LCD 显示器图像的截屏。该数据集由

145000 对重获取图像和自然图像组成。用于重获取图像的设备包括：234 台显示器，173 台扫描仪，282 台打印机和 180 台相机。NTU-Rose[83] 是从 LCD 显示器中重获取得到的高质量图像的集合。原始自然图像共 300 张，使用 3 台数码相机和 3 台液晶显示器重获取 2700 张图像。ICL[85] 数据集也是由 LCD 显示器中重获取得到的图像组成，包括由 9 台不同相机拍摄的 1035 张原始自然图像和使用不同设备获取的 2520 张图像。为了保证重获取图像的图像质量，相机设置进行了调整。因此，该数据集提供了高质量、高分辨率的重获取图像。ASTAR[99] 是一个用于重获取检测的智能手机图像数据集。数据集分为三个子集。子集 A 由 1094 张真实场景图像和 1137 张以真实环境为背景的重获取图像组成。子集 B 是由裁剪子集 A 中的真实场景图像得到的自然图像，以及 1765 张没有真实环境背景的重获取图像构建的。子集 C 由 587 对具有相同内容的自然图像和重获取的图像组成。

8　结论

本文整理了近年来图像源辨识取证领域中的研究，从相机源识别、计算机图形学方法生成图像取证、AI 合成图像取证以及重获取图像取证四个方面分别进行了介绍，研究方法包括传统的基于模型的方法和基于深度学习的方法。虽然图像源辨识取证方面的研究已经取得一定的成果，但目前尚存在一些不足。

（1）图像源辨识取证领域的理论体系还不够完善，大多数的方法还属于经验性或实验性的探索，尚有大量理论问题没有得到解决。例如，学术界对采用什么样的特征信息或者深度网络架构能实现最佳的图像源辨识取证性能还没有形成完善统一的认识。

（2）对不同研究方法的对比缺乏统一开放的比较平台。虽然在图像源辨识取证任务上已有一些公共数据集，但不同研究工作的实验设置不尽相同，如相机源识别中不同方法考虑的相机种类和数量不同，导致实验结果的可比性降低，无法直接对不同研究方法的性能好坏进行对比评价。

（3）对未知图像源辨识的取证方法还处于初期研究阶段。现有的大部分研究方法考虑的检测场景都是在“闭集”上的检测，即训练集中包含的图像源集合包含了测试集中的图像源集合。然而，在现实应用场景中，更常见的是“开集”上的检测，即测试集中涉及的图像来源可能是训练过程中未知的，这对图像源辨识取证技术提出了更严峻的挑战。

（4）图像源辨识取证方法的鲁棒性和通用性需要进一步提升。现有的大多数研究算法框架是基于特定的数据集和特定的取证场景提出的，然而在现实中的检测环境中，需要测试的图像来源广泛、格式多样、内容繁杂、经过不同的后处理操作处理，因此对图像源辨识取证算法的鲁棒性和通用性提出了更高的要求。

综上所述，虽然图像源辨识取证领域的研究已经取得了显著的成果，但在这一领域仍有许多值得探索的研究问题。如何综合考虑图像源辨识取证特征的共性，充分发挥传统方法和深度学习方法的优势，构造高性能、高鲁棒性和高通用性的图像源辨识取证模型，是图像源辨识取证研究成果走向应用的关键。另外，提出一个标准的实验设置协议和共享的数据集，使不同研究方法能够公平地比较，对于这一研究领域的发展也是非常重要的。

参考文献

[1] 周琳娜，王东明. 数字图像取证技术[M]. 北京：北京邮电大学出版社，2008.
ZHOU Linna, WANG Dongming. Forensic technology of digital images[M]. Beijing: Beijing University of Posts and Telecommunications Press, 2008. (in Chinese)

[2] 孙韶杰，吴琼，李国辉. 基于广义高斯分布的图像盲检测算法[J]. 信号处理，2009，25(9)：1388-1392.
SUN Shaojie, WU Qiong, LI Guohui. An image splicing detection approach based on generalized Gaussian distribution and support vector machine[J]. Signal Processing, 2009, 25(9): 1388-1392. (in Chinese)

[3] 赵峰，刘晓腾，荆涛，等. 基于局部块效应的 JPEG 伪造图像的盲取证[J]. 信号处理，2010，26(12)：1805-1811.
ZHAO Feng, LIU Xiaoteng, JING Tao, et al. Blind forensics of JPEG forgeries based on local blocking artifacts [J]. Signal Processing, 2010, 26(12): 1805-1811. (in Chinese)

[4] 卢燕飞，鞠娅莉，于跃，等. 基于图像背景噪声特性的篡改检测[J]. 信号处理，2012，28(9)：1299-1307.

LU Yanfei, JU Yali, YU Yue, et al. Image forgery detection using characteristics of background noise[J]. Signal Processing, 2012, 28(9): 1299-1307. (in Chinese)

[5] 李应灿, 杨建权, 丁峰, 等. 区分来源和目标区域的图像 copy-move 伪造检测方法[J]. 信号处理, 2020, 36(9): 1533-1543.
LI Yingcan, YANG Jianquan, DING Feng, et al. Copy-move detection method for distinguishing between source and target regions[J]. Journal of Signal Processing, 2020, 36(9): 1533-1543. (in Chinese)

[6] CHOI K S, LAM E Y, WONG K K Y. Source camera identification using footprints from lens aberration[C]// Proc SPIE 6069, Digital Photography II, 2006, 6069: 60690J.

[7] LONG Yangjing, HUANG Yizhen. Image based source camera identification using demosaicking[C]// 2006 IEEE Workshop on Multimedia Signal Processing. Victoria, BC, Canada. IEEE, 2006: 419-424.

[8] BAYRAM S, SENCAR H, MEMON N, et al. Source camera identification based on CFA interpolation[C]// IEEE International Conference on Image Processing 2005. Genova, Italy. IEEE, 2005: III-69.

[9] DENG Zhonghai, GIJSENIJ A, ZHANG Jingyuan. Source camera identification using Auto-White Balance approximation[C]//2011 International Conference on Computer Vision. Barcelona, Spain. IEEE, 2011: 57-64.

[10] SORRELL M J. Digital camera source identification through JPEG quantisation[EB/OL]. https://www.igi-global.com/chapter/digital-camera-source-identification-through/26998,2008.

[11] KEE E, JOHNSON M K, FARID H. Digital image authentication from JPEG headers[J]. IEEE Transactions on Information Forensics and Security, 2011, 6(3): 1066-1075.

[12] AL-ANI M, KHELIFI F. On the SPN estimation in image forensics: A systematic empirical evaluation[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(5): 1067-1081.

[13] BONDI L, BAROFFIO L, GÜERA D, et al. First steps toward camera model identification with convolutional neural networks[J]. IEEE Signal Processing Letters, 2017, 24(3): 259-263.

[14] BONDI L, GÜERA D, BAROFFIO L, et al. A preliminary study on convolutional neural networks for camera model identification[J]. Electronic Imaging, 2017, 2017(7): 67-76.

[15] HUANG Na, HE Jingsha, ZHU Nafei, et al. Identification of the source camera of images based on convolutional neural network[J]. Digital Investigation, 2018, 26: 72-80.

[16] YAO Hongwei, QIAO Tong, XU Ming, et al. Robust multi-classifier for camera model identification based on convolution neural network[J]. IEEE Access, 2018, 6: 24973-24982.

[17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770-778.

[18] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 2261-2269.

[19] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 2818-2826.

[20] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 1800-1807.

[21] CHEN Yunshu, HUANG Yue, DING Xinghao. Camera model identification with residual neural network[C]// 2017 IEEE International Conference on Image Processing (ICIP). Beijing, China. IEEE, 2017: 4337-4341.

[22] DING Xinghao, CHEN Yunshu, TANG Zhen, et al. Camera identification based on domain knowledge-driven deep multi-task learning[J]. IEEE Access, 2019, 7: 25878-25890.

[23] RAFI A M, KAMAL U, HOQUE R, et al. Application of DenseNet in camera model identification and post-processing detection[EB/OL]. 2018: arXiv: 1809.00576 [eess.IV]. https://arxiv.org/abs/1809.00576.

[24] FERREIRA A, CHEN Han, LI Bin, et al. An inception-based data-driven ensemble approach to camera model identification[C]//2018 IEEE International Workshop on Information Forensics and Security (WIFS). Hong Kong, China. IEEE, 2018: 1-7.

[25] MARRA F, GRAGNANIELLO D, VERDOLIVA L. On the vulnerability of deep learning to adversarial attacks for

camera model identification[J]. Signal Processing: Image Communication, 2018, 65: 240-248.

[26] TUAMA A, COMBY F, CHAUMONT M. Camera model identification with the use of deep convolutional neural networks[C]//2016 IEEE International Workshop on Information Forensics and Security (WIFS). Abu Dhabi, United Arab Emirates. IEEE, 2016: 1-6.

[27] PEVNY T, BAS P, FRIDRICH J. Steganalysis by subtractive pixel adjacency matrix[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(2): 215-224.

[28] BAYAR B, STAMM M C. Augmented convolutional feature maps for robust CNN-based camera model identification[C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing, China. IEEE, 2017: 4098-4102.

[29] WANG Bo, YIN Jianfeng, TAN Shunquan, et al. Source camera model identification based on convolutional neural networks with local binary patterns coding[J]. Signal Processing: Image Communication, 2018, 68: 162-168.

[30] RAFI A M, TONMOY T I, KAMAL U, et al. RemNet: remnant convolutional neural network for camera model identification[J]. Neural Computing and Applications, 2021, 33(8): 3655-3670.

[31] YANG Pengpeng, NI Rongrong, ZHAO Yao, et al. Source camera identification based on content-adaptive fusion residual networks[J]. Pattern Recognition Letters, 2019, 119: 195-204.

[32] YOU Changhui, ZHENG Hong, GUO Zhongyuan, et al. Multiscale content-independent feature fusion network for source camera identification[J]. Applied Sciences, 2021, 11(15): 6752.

[33] GÜERA D, ZHU Fengqing, YARLAGADDA S K, et al. Reliability map estimation for CNN-based camera model attribution[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, NV, USA. IEEE, 2018: 964-973.

[34] LIU Yunxia, ZOU Zeyu, YANG Yang, et al. Efficient source camera identification with diversity-enhanced patch selection and deep residual prediction[J]. Sensors, 2021, 21(14): 4701.

[35] BAYAR B, STAMM M C. Towards open set camera model identification using a deep learning framework[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB, Canada. IEEE, 2018: 2007-2011.

[36] MAYER O, STAMM M C. Learned forensic source similarity for unknown camera models[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB, Canada. IEEE, 2018: 2012-2016.

[37] NG T T, CHANG S F, HSU J, et al. Physics-motivated features for distinguishing photographic images and computer graphics[C]//Proceedings of the 13th annual ACM international conference on Multimedia. Hilton Singapore. ACM, 2005: 239-248.

[38] CHEN Wen, SHI Y Q, XUAN Guorong. Identifying computer graphics using HSV color model and statistical moments of characteristic functions[C]//2007 IEEE International Conference on Multimedia and Expo. Beijing, China. IEEE, 2007: 1123-1126.

[39] GALLAGHER A C, CHEN T. Image authentication by detecting traces of demosaicing[C]//2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Anchorage, AK, USA. IEEE, 2008: 1-8.

[40] LI Zhaohong, YE Jingyu, SHI Yunqing. Distinguishing computer graphics from photographic images using local binary patterns[C]//The International Workshop on Digital Forensics and Watermarking 2012, 2013: 228-241. DOI:10.1007/978-3-642-40099-5_19.

[41] LYU S, FARID H. How realistic is photorealistic?[J]. IEEE Transactions on Signal Processing, 2005, 53(2): 845-850.

[42] SANKAR G, ZHAO V, YANG Y H. Feature based classification of computer graphics and real images[C]//2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Taipei, Taiwan, China. IEEE, 2009: 1513-1516.

[43] IANEVA T I, DE VRIES A P, ROHRIG H. Detecting cartoons: A case study in automatic video-genre classification[C]//2003 International Conference on Multimedia and Expo. ICME'03. Proceedings (Cat. No.03TH8698). Baltimore, MD, USA. IEEE, 2003: I-449.

[44] ZHANG Rong, WANG Rangding, NG T T. Distinguishing photographic images and photorealistic computer graphics using visual vocabulary on local image edges[C]//Digital Forensics and Watermarking, 2012: 292-305. DOI:10.1007/978-3-642-32205-1_24.

[45] PENG Fei, ZHOU Dielan, LONG Min, et al. Discrimi-

nation of natural images and computer generated graphics based on multi-fractal and regression analysis[J]. AEU-International Journal of Electronics and Communications, 2017, 71: 72-81.

[46] SUTTHIWAN P, YE Jingyu, SHI Y Q. An enhanced statistical approach to identifying photorealistic images[M]// Digital Watermarking. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009: 323-335.

[47] YU I J, KIM D G, PARK J S, et al. Identifying photorealistic computer graphics using convolutional neural networks[C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing, China. IEEE, 2017: 4093-4097.

[48] HE Ming. Distinguish computer generated and digital images: A CNN solution[J]. Concurrency and Computation: Practice and Experience, 2019, 31(12): e4788.

[49] HE Peisong, JIANG Xinghao, SUN Tanfeng, et al. Computer graphics identification combining convolutional and recurrent neural networks[J]. IEEE Signal Processing Letters, 2018, 25(9): 1369-1373.

[50] YAO Ye, HU Weitong, ZHANG Wei, et al. Distinguishing computer-generated graphics from natural images based on sensor pattern noise and deep learning[J]. Sensors, 2018, 18(4): 1296.

[51] QUAN Weize, WANG Kai, YAN Dongming, et al. Distinguishing between natural and computer-generated images using convolutional neural networks[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2772-2787.

[52] RAHMOUNI N, NOZICK V, YAMAGISHI J, et al. Distinguishing computer graphics from natural images using convolution neural networks[C]//2017 IEEE Workshop on Information Forensics and Security (WIFS). Rennes, France. IEEE, 2017: 1-6.

[53] NGUYEN H H, TIEU T N D, NGUYEN-SON H Q, et al. Modular convolutional neural network for discriminating between computer-generated images and photographic images[C]//Proceedings of the 13th International Conference on Availability, Reliability and Security. Hamburg Germany. New York, NY, USA: ACM, 2018: 1-10.

[54] DE REZENDE E R S, RUPPERT G C S, THEÓPHILO A, et al. Exposing computer generated images by using deep convolutional neural networks[J]. Signal Processing: Image Communication, 2018, 66: 113-126.

[55] ZHANG Ruisong, QUAN Weize, FAN Lubin, et al. Distinguishing computer-generated images from natural images using channel and pixel correlation[J]. Journal of Computer Science and Technology, 2020, 35(3): 592-602.

[56] HE Peisong, LI Haoliang, WANG Hongxia, et al. Detection of computer graphics using attention-based dual-branch convolutional neural network from fused color components[J]. Sensors, 2020, 20(17): 4743.

[57] TOLOSANA R, VERA-RODRIGUEZ R, FIERREZ J, et al. Deepfakes and beyond: A Survey of face manipulation and fake detection[J]. Information Fusion, 2020, 64: 131-148.

[58] VERDOLIVA L. Media forensics and DeepFakes: An overview[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(5): 910-932.

[59] KARRAS T, AILA, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation[EB/OL]. https://openreview.net/pdf? id=Hk99zCeAb,2017.

[60] KARRAS T, LAINE S, AILA Timo. A style-based generator architecture for generative adversarial networks[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 4396-4405.

[61] ZHU Junyan, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2242-2251.

[62] CHOI Y, CHOI M, KIM M, et al. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 8789-8797.

[63] ZHANG Han, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks[C]//International Conference on Machine Learning. Long Beach, California. PMLR, 2019, 97:7354-7363.

[64] BROCK A, DONAHUE J, SIMONYAN K. Large scale GAN training for high fidelity natural image synthesis[C]//Proceedings of International Conference on Learning Representations(ICLR). New Orleans, Louisiana, United States, 2019.

[65] MCCLOSKEY S, ALBRIGHT M. Detecting GAN-generated imagery using saturation cues[C]//2019 IEEE International Conference on Image Processing (ICIP). Taipei,

Taiwan, China. IEEE, 2019: 4584-4588.

[66] ZHANG X, KARAMAN S, CHANG S F. Detecting and simulating artifacts in GAN fake images[C]//2019 IEEE International Workshop on Information Forensics and Security (WIFS). Delft, Netherlands. IEEE, 2019: 1-6.

[67] LI Haodong, LI Bin, TAN Shunquan, et al. Identification of deep network generated images using disparities in color components[J]. Signal Processing, 2020, 174: 107616.

[68] MARRA F, GRAGNANIELLO D, COZZOLINO D, et al. Detection of GAN-generated fake images over social networks[C]//2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). Miami, FL, USA. IEEE, 2018: 384-389.

[69] FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.

[70] COZZOLINO D, POGGI G, VERDOLIVA L. Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection[C]// Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security. Philadelphia Pennsylvania USA. New York, NY, USA: ACM, 2017: 159-164.

[71] BAYAR B, STAMM M C. A deep learning approach to universal image manipulation detection using a new convolutional layer[C]//IH&MMSec'16: Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, 2016: 5-10.

[72] LI Haodong, CHEN Han, LI Bin, et al. Can forensic detectors identify GAN generated images? [C]//2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). Honolulu, HI, USA. IEEE, 2018: 722-727.

[73] MO Huaxiao, CHEN Bolin, LUO Weiqi. Fake faces identification via convolutional neural network[C]//Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. Innsbruck Austria. New York, NY, USA: ACM, 2018: 43-47.

[74] NATARAJ L, MOHAMMED T M, MANJUNATH B S, et al. Detecting GAN generated fake images using co-occurrence matrices[J]. Electronic Imaging, 2019, 2019(5): 532-1.

[75] WANG Shengyu, WANG O, ZHANG R, et al. CNN-generated images are surprisingly easy to spot… for now [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 8692-8701.

[76] ZHUANG Yixiu, HSU C C. Detecting generated image based on a coupled network with two-step pairwise learning [C]//2019 IEEE International Conference on Image Processing (ICIP). Taipei, Taiwan, China. IEEE, 2019: 3212-3216.

[77] MI Zhongjie, JIANG Xinghao, SUN Tanfeng, et al. GAN-generated image detection with self-attention mechanism against GAN generator defect[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(5): 969-981.

[78] FRANK J, EISENHOFER T, SCHÖNHERR L, et al. Leveraging frequency analysis for deep fake image recognition[EB/OL]. https://deepfakechallenge.com/Download/Docs/1539-Paper.pdf,2020.

[79] AGARWAL S, GIRDHAR N, RAGHAV H. A novel neural model based framework for detection of GAN generated fake images[C]//2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). Noida, India. IEEE, 2021: 46-51.

[80] YU Hang, NG T T, SUN Qibin. Recaptured photo detection using specularity distribution[C]//2008 15th IEEE International Conference on Image Processing. San Diego, CA, USA. IEEE, 2008: 3140-3143.

[81] GAO Xinting, NG T T, QIU Bo, et al. Single-view recaptured image detection based on physics-based features [C]//2010 IEEE International Conference on Multimedia and Expo. Singapore. IEEE, 2010: 1469-1474.

[82] YIN Jing, FANG Yanmei. Markov-based image forensics for photographic copying from printed picture[C]//Proceedings of the 20th ACM international conference on Multimedia-MM'12. Nara, Japan. New York: ACM Press, 2012: 1113-1116.

[83] CAO Hong, KOT A C. Identification of recaptured photographs on LCD screens[C]//2010 IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, TX, USA. IEEE, 2010: 1790-1793.

[84] YIN Jing, FANG Yanmei. Digital image forensics for photographic copying[C]//Media Watermarking, Security, and Forensics 2012. Burlingame, California, USA. SPIE, 2012: 83030F1-83030F7.

[85] THONGKAMWITOON T, MUAMMAR H, DRAGOTTI P L. An image recapture detection algorithm based on learning dictionaries of edge profiles[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(5): 953-968.

[86] ANJUM A, ISLAM S. Recapture detection technique based on edge-types by analysing high-frequency components in digital images acquired through LCD screens [J]. Multimedia Tools and Applications, 2020, 79(11/12): 6965-6985.

[87] LI R, NI R, ZHAO Y. An effective detection method based on physical traits of recaptured images on LCD screens[C]// International Workshop on Digital Watermarking. Tokyo, Japan. Springer, 2015: 107-116.

[88] SUN Yanjun, SHEN Xuanjing, LIU Changming, et al. Recaptured image forensics algorithm based on image texture feature[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(3): 2054011.

[89] YANG Pengpeng, NI Rongrong, ZHAO Yao. Recapture image forensics based on Laplacian convolutional neural networks [M] // Digital Forensics and Watermarking. Cham: Springer International Publishing, 2017: 119-128.

[90] CHOI H Y, JANG H U, SON J, et al. Content recapture detection based on convolutional neural networks[C]// Information Science and Applications 2017, 2017: 339-346. DOI:10.1007/978-981-10-4154-9_40.

[91] LI Haoliang, WANG Shiqi, KOT A. Image recapture detection with convolutional and recurrent neural networks [J]. Electronic Imaging, 2017, 2017(7): 87-91.

[92] AGARWAL S, FAN Wei, FARID H. A diverse large-scale dataset for evaluating rebroadcast attacks[C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, AB, Canada. IEEE, 2018: 1997-2001.

[93] GLOE T, BOHME R. The Dresden image database for benchmarking digital image forensics[J]. Journal of Digital Forensic Practice, 2010, 3(2/4): 150-159.

[94] SHULLANI D, FONTANI M, IULIANI M, et al. VISION: a video and image dataset for source identification [J]. EURASIP Journal on Information Security, 2017, 2017(1): 15.

[95] IEEE Signal Processing Society. IEEE signal processing cup 2018: Forensic camera model identification challenge. https://signalprocessingsociety.org/getinvolved/signal-processing-cup.

[96] NG T T, CHANG S F, HSU J, et al. Columbia photographic images and photorealistic computer graphics dataset [EB/OL]. https://www.ee.columbia.edu/ln/dvmm/publications/05/ng_cgdataset_05.pdf,2005.

[97] TOKUDA E, PEDRINI H, ROCHA A. Computer generated images vs. digital photographs: A synergetic feature and classifier combination approach[J]. Journal of Visual Communication and Image Representation, 2013, 24(8): 1276-1292.

[98] GUAN Haiying, KOZAK M, ROBERTSON E, et al. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation [C] // 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). Waikoloa, HI, USA. IEEE, 2019: 63-72.

[99] GAO Xinting, QIU Bo, SHEN Jingjing, et al. A smart phone image database for single image recapture detection [M]//Digital Watermarking. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 90-104.

作者简介

陈艺芳 女,1990年生,云南昆明人。广东技术师范大学网络空间安全学院校聘副教授,硕士生导师,研究方向为多媒体信息取证。

E-mail: chenyf@gpnu.edu.cn

何自强 男,1998年生,江西南昌人。中山大学计算机学院硕士研究生,主要研究方向为深度篡改检测。

E-mail: hezq7@mail2.sysu.edu.cn

文冠臣 男,1999年生,河北邢台人。广东技术师范大学网络空间安全学院硕士研究生,主要研究方向为图像内容安全。

E-mail: 1057968435@qq.com

康显桂(通信作者) 中山大学数据科学与计算机学院教授/博士生导师,兼任广东省信息安全重点实验室副主任,广东省优秀博士论文奖获得者,IEEE期刊编委,亚太信息与信号处理协会(APSIPA)信息取证与安全技术委员会的主席,入榜“中国高被引学者”。从事信息取证、信息隐藏等多媒体信息安全方面的研究。

E-mail: isskxg@mail.sysu.edu.cn