

引用格式:方梦华,林果园. 基于孪生网络的数字图像相机源识别[J]. 微电子学与计算机,2021,38(6):82-86. [FANG M H, LIN G Y. Source identification of digital image based on Siamese network[J]. Microelectronics & Computer, 2021, 38(6): 82-86.]

基于孪生网络的数字图像相机源识别

方梦华^{1,2}, 林果园^{1,2}

(1 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116; 2 矿山数字化教育部工程研究中心, 江苏 徐州 221116)

摘 要: 传统的利用传感器模式噪声的相机源识别方法, 自动化程度低. 直接使用深度学习方法进行识别, 提取特征阶段又更倾向于学习到图像内容有关的特征, 而不是相机“指纹”特征. 为了解决上述问题, 本文设计了一种基于孪生网络的数字图像相机源识别架构, 两个网络分支共享权重. 采用加入注意力机制的深度残差网络进行特征提取, 抑制图像内容等因素对整个相机源识别任务的影响. 对比实验结果表明, 本文提出的方法在一定程度上提高了相机源识别的准确率.

关键词: 图像来源鉴别; 相机源识别; 孪生网络; 传感器模式噪声; 注意力机制

中图分类号: TP391.4

文献标识码: A

文章编号: 1000-7180(2021)06-0082-05

Source identification of digital image based on Siamese network

FANG Meng-hua^{1,2}, LIN Guo-yuan^{1,2}

(1 School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, Jiangsu China; 2 Engineering Research Center of the Ministry of Education for the Digitization of Mines, Xuzhou 221116, Jiangsu China)

Abstract: The traditional camera source identification method using sensor pattern noise has a low degree of automation. The deep learning method is directly used for identification, and the feature extraction stage is more inclined to learn the features related to the image content, rather than the camera "fingerprint" feature. In order to solve the above problems, this paper designs a digital image camera source identification architecture based on the Siamese network, and the two network branches share the weight. A deep residual network with an attention mechanism is used for feature extraction to suppress the influence of image content on the entire camera source identification task. The comparative experiment results show that the method proposed in this paper improves the accuracy of camera source identification to a certain extent.

Key words: image source identification; camera source identification; siamese network; sensor pattern noise; attention mechanism

1 引言

数字图像的来源识别是图像取证的一个重要组成部分^[1]. 数字图像的来源鉴别, 即成像设备识别, 是指给定一张图像, 通过科技手段和方法, 判定它的成像设备是什么. 来源鉴别的种类有很多, 本文主要

研究的是个体识别相关的内容, 给定一张数字图像, 判断它是来自几个相机中的哪一个. 要实现这项任务, 传统方法是通过传感器模式噪声来实现^[2]. 传感器模式噪声的提出最早是在 2006 年, Lucas 等人在文献[3]中提出了最早的基于传感器模式噪声的相机源识别算法. 之后, Chen 等人在文献[4]中对该

收稿日期: 2020-08-03; 修回日期: 2020-09-06

基金项目: 软件新技术国家重点实验室开放基金项目(KFKT2018B27); 中央高校基础研究基金项目(2017XKQY079)

算法进行改进,结合图像光响应非均匀性进行检测.国内的陈凌翔等人^[5]通过去除颜色插值所产生的一部分噪声,提高模式噪声的提取精度.上述研究主要是依靠手动完成的.此后出现了不少使用深度学习进行相机源的方法^[6-8],但是直接使用深度学习方法容易受到数据集的影响.它更倾向于学习图像内容特征,容易使相机源识别任务失去方向.一旦样本数量较少或者质量不高,就会达不到预期效果.甚至文献^[9]指出,在一定条件下,深度学习方法进行相机源识别不如基于传感器模式噪声的效果好.因此,本文设计了一种基于孪生网络的识别架构来进行相机源鉴别.将基于传感器模式噪声的方法和深度学习方法进行结合,在孪生网络中首先提取相机的“指纹”,之后通过添加了注意力机制的残差网络提取特征,实现相机源识别.

2 基于孪生网络的相机源识别架构

2.1 孪生网络整体架构

为了实现相机源识别,本文设计了一种基于孪生网络的识别架构.整体架构如图1所示.虚线框里是孪生网络共享权重的两个分支网络.在设计的孪生网络中,需要输入两个样本,即基准图像 P_1 和待测图像 P_2 .之后对输入的两个图像进行滤波操作,从而得到相机参考模式噪声 S_1 和待测图像模式噪声 S_2 .由基准图像 P_1 提取到的模式噪声即为相机C的“指纹”信息.利用添加了注意力机制的残差网络,对两个噪声图像提取特征.通过计算特征向量之间的距离,将待测图像的噪声信息与相机C的“指纹”信息进行相似度度量,从而判断待测图像 P_2 是否可以看成是相机C拍摄的,实现相机源判别.

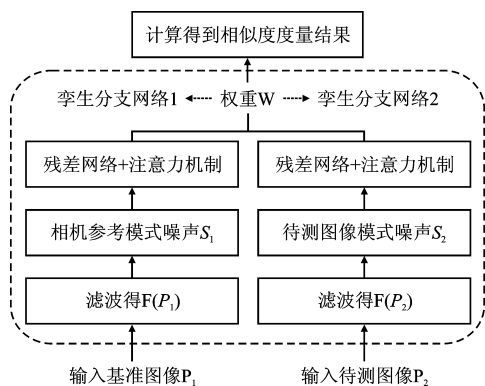


图1 孪生网络整体架构

在进行相机源识别之前对孪生网络进行训练.训练阶段每次输入一对图像以及这对图像是否来自同一个相机的标签.整个孪生网络利用损失函数来

进行优化,通过梯度反传更新两个分支网络的共享权重.损失函数如公式(1)和公式(2)所示.

$$L(W(Y, X_1, X_2)) = \frac{1}{2N} \sum_{n=1}^N Y D_w^2 + (1-Y) \max(m - D_w, 0)^2 \quad (1)$$

$$D_w(X_1, X_2) = \|X_1 - X_2\|_2 = \left(\sum_{i=1}^P (X_1^i - X_2^i)^2 \right)^{\frac{1}{2}} \quad (2)$$

式中, D_w 代表提取到的一个图像对中,特征向量的欧氏距离(二范数), P 表示提取到的特征维数; Y 是表示两个图像是否来自同一相机的标签. $Y=1$ 代表输入的一对图像来自同一个相机, $Y=0$ 则代表两者不是同一相机拍摄; m 为设定的阈值; N 为样本个数.如果输入的两个图像是同一相机拍摄的,那么损失函数就剩下 $\sum Y D_w^2$,此时如果特征向量的欧式距离较大,那么损失函数也随之变大,相似的特征离得较远,说明当前的网络模型仍需要优化,不断更新匹配程度.如果输入的两个图像不是同一相机拍摄的,损失函数为 $\sum (1-Y) \max(m - D_w, 0)^2$,此时若欧式距离小的话,损失函数会变大,不相似的特征离得较近,说明网络模型还没有达到最佳.其中阈值 m 的设置是为了表示特征向量的欧式距离只考虑在 $0 \sim m$ 之间的.阈值 m 是根据样本特征向量的距离与损失函数值之间的关系进行设置的.对于来自不同相机的样本对,损失函数值随着特征向量距离的增大而减小,损失函数值降为零时,特征向量之间的欧式距离即为被看作临界值,也就是阈值 m .欧式距离大于阈值 m 时,损失函数看作0.

2.2 数字图像的“指纹”提取

每个成像设备都有其独一无二的传感器.自然景物在成像过程中,光信号传到相机的传感器,此时由于各像素对光的敏感程度是有一定差异的,同时还有低频缺陷的影响,所以每个传感器都有其独一无二的模式噪声,相当于相机独特的“指纹”.换句话说,所有由该相机拍摄的图像,模式噪声都带有该相机的痕迹,基于此可进行相机源的判别.

本文在孪生网络前端进行相机“指纹”的提取,通过对图像进行自适应小波滤波来实现.原图像跟滤波之后的图像相减,得到的差值噪声即为所求.为了判断待测图像 P_2 是否是指定相机C拍摄,需要对比待测图像提取到的“指纹”信息和相机C的“指纹”信息.待测图像的“指纹”也就是传感器模式噪声可以直接提取.相机的参考模式噪声不能直接提取,需要借助由该相机拍摄的参考图像进行提取,主要

依据是所有由该相机拍摄的数字图像都该相机的痕迹. 理想情况下, 所有参考图像的传感器模式噪声是相同的. 也就是说, 任意一张参考图像的传感器模式噪声都可以代表相机的“指纹”. 但是在实际的“指纹”提取中, 不管用何种滤波方式, 得到的传感器模式噪声都会受到图像内容以及其他噪声等的影响. 为了更有效地进行相机源识别, 将图像内容对传感器模式噪声的影响降到最低, 本文通过选取图像内容简单的参考图像来进行相机源识别. 首先对相机

C 有限的参考图像计算图像复杂度, 具体的图像复杂度计算参考文献[10]中的方法. 按照计算得到的图像复杂度, 将参考图像从低到高排序, 挑选复杂度最低的参考图像, 就是基准图像 P_1 , 作为孪生网络的其中一个输入. 选定基准图像 P_1 , 对其进行滤波得到的差值噪声, 即可作为相机 C 的“指纹”. 通过和待测图像的传感器模式噪声进行相似度度量, 完成相机源识别任务.

2.3 孪生网络分支设计

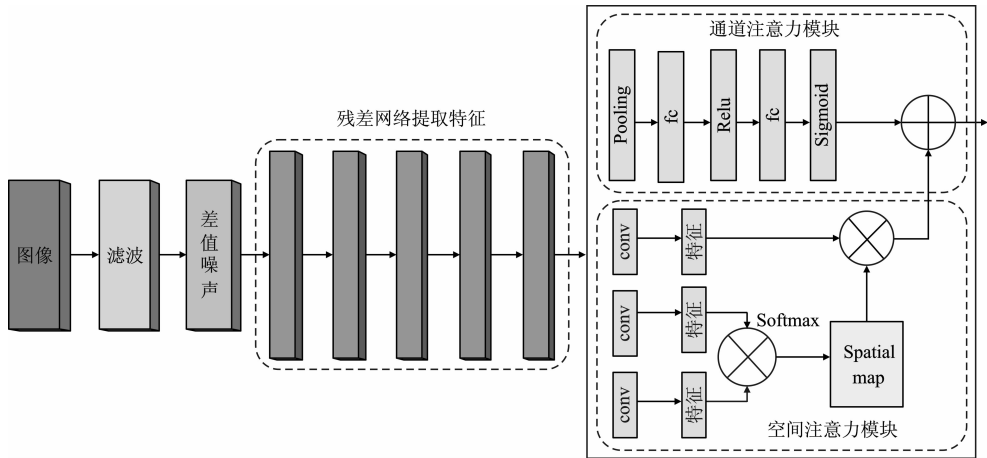


图 2 分支网络结构

孪生的分支网络具体结构如图 2 所示. 考虑到要提取深层次的特征, 同时又要避免网络层次多造成网络退化的情况, 采用了残差网络, 网络结构如表 1 所示. 之所以利用残差网络是因为设计的网络中层数增多时, 容易出现性能下降的情况, 而残差网络可以很好地解决这个问题. 深层次的残差网络也可以有效提取到模式噪声更细微的特征.

五个部分组成, 具体结构如表 1 所示.

本文采用的残差网络由一系列残差单元组成, 单个残差单元结构如图 3 所示. 输入 X , 通过残差单元, 输出的是 $H(x) = F(x) + X$. 与普通的卷积网络不同, 普通卷积网络输出的是 $F(x)$, 根本原因是残差网络结构中有两种映射. 其中一种是本身映射, 指的是图 3 中 identity 的折线, 另一种残差映射, 指的是除了折线的两个卷积层, 所以最后的输出是 $H(x) = F(x) + X$. 残差中的“差”指的就是 $F(x)$ 的部分. 在前向过程中, 当浅层的输出已经足够成熟, 让深层网络后面的层能够实现恒等映射的作用. 如果深层网络的后面那些层是恒等映射, 那么模型就退化为一个浅层网络. 换句话说就是, 残差网络跟普通网络相比, 增加了一些可以跳过某些层的快捷连接, 将数据传送到更深层, 一定程度上保证了信息的完整性. 在此基础上, 本文设计了分支网络中使用的残差网络结构, 主要由

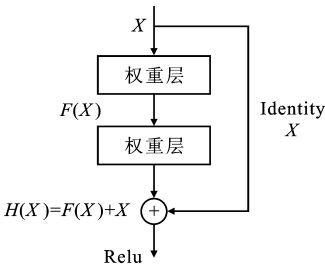


图 3 残差单元结构

表 1 分支网络中的残差网络结构表

Layer_name	Parameter	Output_size
conv_1	Kernel_size 7×7	112×112
	Num_out 64 Stride 2	
conv_2	$\begin{bmatrix} 3 \times 3 & 64 \\ 3 \times 3 & 64 \end{bmatrix} \times 3$	56×56
conv_3	$\begin{bmatrix} 3 \times 3 & 128 \\ 3 \times 3 & 128 \end{bmatrix} \times 4$	28×28
conv_4	$\begin{bmatrix} 3 \times 3 & 256 \\ 3 \times 3 & 256 \end{bmatrix} \times 6$	14×14
conv_5	$\begin{bmatrix} 3 \times 3 & 512 \\ 3 \times 3 & 512 \end{bmatrix} \times 3$	7×7

注意力机制可以从大量信息中有选择地筛选出少量重要信息, 并聚焦到这些重要信息上, 忽略大多

不重要的信息.在残差网络后加入通道注意力机制来提取有效特征的原理是,卷积神经网络输入的图片一般会通过 (R,G,B) 三个通道表示出来,通过卷积层的卷积操作后,每一个通道就又会生成新的信号.通道注意力机制主要通过给每个通道的新信息添加权重,以此实现有效特征的提取.权重越大,说明该通道的信息对我们想要注意的目标越重要.除了通道注意力机制,加入的空间注意力机制在保留原始图像相关重要信息的情况下,可以将初始图像的空间信息提取到另一个空间模块中.一般情况下,卷积神经网络的池化层利用平均池化等方法,压缩提取到的图像信息,实现减少运算量的目的.但是这样直接进行池化操作可能导致我们所需要的关键信息无法提取到.因此,使用空间注意力机制,有利于将我们需要的关键信息保留下来.

两个注意力机制在相机源识别任务中的具体应用如图 2 所示.对于孪生网络中添加的通道注意力机制,输入特征 $R \in M^{w \times h \times d}$,通过平均池化层得到特征向量 $r=[r_1,r_2,\cdots,r_n]$, n 是通道数量.然后依次通过全连接层、Relu 激活层、全连接层和 Sigmoid 激活层,输出向量 $\alpha=[\alpha_1,\alpha_2,\cdots,\alpha_n]$.根据公式(3)得到通道注意力特征图 R^1 .在空间注意力机制中,输入特征 $B \in N^{w \times h \times d}$,经过三个卷积层得到特征图 O,P,Q ,将特征图 P,Q 转换为 $N^{w \times h \times d}$.利用公式(4)得到空间注意力图,其中, $S_{i,j}$ 表示第 i 与 j 两个区域之间的权重.此时将特征图 O 也转换为 $N^{w \times h \times d}$,利用公式(5)得到空间注意力特征图 B^1 .

$$\alpha \cdot R \rightarrow R^1, R^1 \in M^{w \times h \times d}$$

(3)

$$S_{i,j} = \frac{\exp(P_i^T \cdot Q_j)}{\sum_{i=1}^{wh} \exp(P_i^T \cdot Q_j)}$$

(4)

$$O \cdot S^T \rightarrow B^1, B^1 \in N^{w \times h \times d}$$

(5)

通道注意力机制和空间注意力机制本质上就是,根据各通道和各空间信息对于相机源识别任务的贡献不同,对它们分配不同的权重,从而达到加强有用信息、抑制无用信息的目的.因此,获得通道注意力特征和空间注意力特征之后,直接将两者的元素进行相加,就得到我们需要的注意力机制特征.如公式 6 所示.以此来对模式噪声进行更好的特征表征.

表 3 传统的相机源识别方法和本文提出的方法的准确率对比 (%)

Method	Casio-EX Z150	FujiFilm FinePixJ50	Kodak M1063	Nikon D200	Olympus mju	Panasonic DMC	Pentax OptioA40
Basic SPN	98	85	91	95	93	94	95
AlexNet	52	56	67	73	62	50	65
Proposed	98	95	95	98	99	97	98

$$Z_i = R_i^1 + B_i^1, i = 1, 2, \cdots, n$$

(6)

3 实验方法和结果分析

3.1 实验方法和设置

为了验证所提出相机源识别方法的有效性,对设计的基于孪生网络架构的相机源识别方法设置了对比实验.实验中用来进行参照对比的其他两种方法分别是,传统的基于传感器模式噪声的识别方法^[3],以及直接使用深度学习 AlexNet 网络进行相机源识别的方法.通过与这两种方法进行对比,验证本文提出的方法在操作便捷性和识别准确率方面的优势.

实验中数据集采用 Dresden 公开图像数据集^[11],选取了来自其中 7 个摄像机的 1 050 张图片.表 2 展示了实验中使用的七个数码相机的相机型号等信息.实验使用 Tensorflow 框架,编程语言使用 python.实验的训练集,验证集和测试集按照 8:1:1 的比例分配.实验采用 Adagrad 自适应梯度优化算法,学习率设置为 0.01,批大小为 128,迭代次数为 900.为了得到可靠的实验结果,实验结果是通过随机进行 3 次试验平均得到的.

表 2 实验所用数据集的细节信息

相机编号	相机型号	图像数量
1	Casio-EX-Z150	150
2	FujiFilm-FinePixJ50	150
3	Kodak-M1063	150
4	Nikon-D200	150
5	Olympus-mju	150
6	Panasonic-DMC	150
7	Pentax-OptioA40	150

3.2 实验结果及分析

为了方便比较,本文将不同分辨率的测试图像统一成尺寸为 224×224 的图像块来验证所设计的方法的性能.采用识别准确率和受试者工作特性曲线 ROC 作为评价指标来衡量图像源识别的性能.准确率定义为所有正确识别的图像数量与所有识别的图像数量之比.表 3 是本文提出的方法和基准方法的识别准确率对比.

从表 3 中可以得到,利用基础的传感器模式噪声识别方法进行相机源鉴别时,平均准确率为 93%。采用计算机视觉领域常用的 AlexNet 作为孪生网络的分支网络,进行相机源识别时,识别准确率普遍较低。这是因为该方法受到数据集样本数量的影响较大,同时,该网络更倾向于学习到图像内容相关的特征。这样的结果也在一定程度上证明了,使用不加预处理的神经网络进行相机源识别,可能不如基于传感器模式噪声进行相机源识别的效果好。采用本文设计的方法进行实验,平均准确率达到 97%,比基于传感器模式噪声识别方法的平均准确率提高了 4%。这说明,本文设计的识别方法不仅比传统方法更便捷、自动化程度高,而且识别率也有一定的提高。

除了准确率指标外,实验还采用 ROC 用于表征相机模型的识别性能。通过绘制各种阈值设置下的真阳性率 (TPR) 与误报率 (FPR) 可以创建 ROC,结果如图 4 所示。横坐标 FPR 代表负例被错误分为正例的概率,如公式 7 所示。其中,FP 表示负类被预测为正类的数量,TN 表示负类被预测为负类的数量。纵坐标 TPR 代表将正例分对的概率,如公式 8 所示。其中,TP 表示正类被预测为正类的数量,FN 表示正类被预测为负类的数量。

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

$$TPR = \frac{TP}{TP + FN} \quad (8)$$

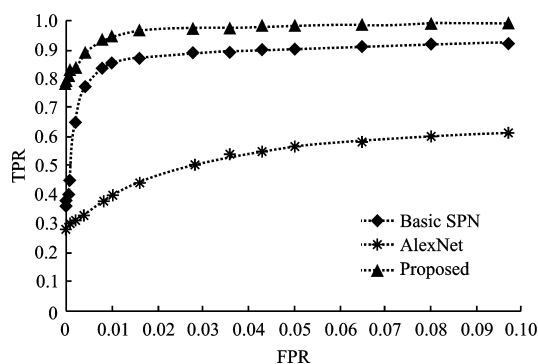


图 4 ROC 曲线比较

由图 4 可以看出,当 TPR 越大,FPR 越小时,分类效果越好。三种相机源识别方法进行对比,基础的传感器模式噪声识别方法分类效果好于基于 AlexNet 网络的相机源识别方法。同时,本文提出的基于孪生网络的识别架构,实验显示分类效果比其他两种方法都好。这是因为其不仅在孪生网络的分支网络采用了加入注意力机制的深度残差网络,同时还在网络前端对输入网络的数字图像进行预处理,利用处理得到的模式噪声进行源鉴别,相当于将

传感器模式噪声方法和深度学习方法的优点进行了融合,故分类效果比其他两种方法好。

4 结束语

传统的利用传感器模式噪声的相机源识别方法,识别过程繁琐、自动化程度低,而直接使用深度学习的相机源识别方法更倾向于学习到图像内容相关的特征,且对数据集的样本数量和质量依赖较大。针对这样的问题,本文将传感器模式噪声的识别方法和深度学习有效结合起来,提出了一种基于孪生网络的相机源识别方法。孪生网络分支前端首先对输入的图像进行滤波,得到传感器模式噪声作为相机的指纹。然后利用加入注意力机制的深度残差网络对其进行特征提取。最后通过孪生网络的相似度度量判断识别结果。对比实验结果表明,我们的方法在识别准确率上面占有一定的优势。未来可以在以下几方面继续改进:可以尝试三胞胎网络等其他网络架构。关于分支网络的设计可以尝试其他更深层的网络结构。同时,未来在相机源识别的实际应用中,实时性问题也要进一步考虑。

参考文献:

- [1] 蒋翔, 韦世奎, 赵瑞珍, 等. 数字图像的设备溯源技术综述[J]. 北京交通大学学报, 2019, 43(2): 48-57. DOI: 10.11860/j.issn.1673-0291.20190036. JIANG X, WEI S K, ZHAO R Z, et al. Survey of imaging device source identification[J]. Journal of Beijing Jiaotong University, 2019, 43(2): 48-57. DOI: 10.11860/j.issn.1673-0291.20190036.
- [2] 蒋英春. 基于成像传感器特征的相机源识别[J]. 计算机工程与应用, 2012, 48(18): 193-196. DOI: 10.3778/j.issn.1002-8331.2012.18.041. JIANG Y C. Source camera identification based on characteristics of imaging sensor[J]. Computer Engineering and Applications, 2012, 48(18): 193-196. DOI: 10.3778/j.issn.1002-8331.2012.18.041.
- [3] LUKASJ, FRIDRICH J, GOLJAN M. Digital camera identification from sensor pattern noise[J]. IEEE Transactions on Information Forensics and Security, 2006, 1(2): 205-214. DOI: 10.1109/TIFS.2006.873602.
- [4] CHEN M, FRIDRICH J, GOLJAN M, et al. Determining image origin and integrity using sensor noise[J]. IEEE Transactions on Information Forensics and Security, 2008, 3(1): 74-90. DOI: 10.1109/TIFS.2007.916285.

issn1000-7180, 2019, 04, 012.

- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2015.
 - [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6517-6525. DOI: 10.1109/CVPR.2017.690.
-
- (上接第 86 页)
- [5] 陈凌翔, 柯永振, 孙玉婷. 一种用于图片来源检测的相机参考模式噪声提取算法[J]. 计算机工程, 2017, 43(12): 227-230. DOI: 10.3969/j.issn.1000-3428.2017.12.041.
CHEN L X, KE Y Z, SUN Y T. A camera reference pattern noise extraction algorithm for image source detection[J]. Computer Engineering, 2017, 43(12): 227-230. DOI: 10.3969/j.issn.1000-3428.2017.12.041.
 - [6] FERREIRA A, CHEN H, LI B, et al. An inception-based data-driven ensemble approach to camera model identification[C]//Proceedings of 2018 IEEE International Workshop on Information Forensics and Security. Hong Kong, China: IEEE, 2018: 1-7. DOI: 10.1109/WIFS.2018.8630774.
 - [7] MAYER O, STAMM M C. Learned forensic source similarity for unknown camera models[C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada: IEEE, 2018: 2012-2016. DOI: 10.1109/ICASSP.2018.8462585.
 - [8] ZHAO M N, WANG B, WEI F, et al. Source camera identification based on coupling coding and adaptive filter[J]. IEEEAccess, 2020, 8: 54431-54440. DOI: 10.1109/ACCESS.2019.2959627.

作者简介:

程钰清 女, (1995-), 硕士研究生. 研究方向为计算机视觉.
贺占庄(通讯作者) 男, (1962-), 研究员. 研究方向为航天嵌入式系统结构设计. E-mail: hzz771@163.com
马 钟 男, (1971-), 博士, 高级工程师. 研究方向为计算机视觉.
毕瑞星 男, (1989-), 硕士, 工程师. 研究方向为计算机视觉.
毛远宏 男, (1980-), 博士研究生. 研究方向为计算机视觉.

- [9] AHMED F, KHELIFI F, LAWGALY A, et al. Comparative analysis of a deep convolutional neural network for source camera identification[C]//Proceedings of the 2019 IEEE 12th International Conference on Global Security, Safety and Sustainability. London: IEEE, 2019: 1-6. DOI: 10.1109/ICGS3.2019.8688260.
- [10] AN Y, WANG B, ZHAO M J, et al. Patch-based sensor pattern noise for camera source identification [C]//Proceedings of 2015 IEEE China Summit and International Conference on Signal and Information Processing. Chengdu: IEEE, 2015: 866-870. DOI: 10.1109/ChinaSIP.2015.7230528.
- [11] GLOE T, BÖHME R. The 'Dresden Image Database' for benchmarking digital image forensics [C]//Proceedings of the 2010 ACM Symposium on Applied Computing. Sierre, Switzerland: ACM, 2010: 1584-1590. DOI: 10.1145/1774088.1774427.

作者简介:

方梦华 女, (1995-), 硕士研究生. 研究方向为图像取证、云计算与信息安全. E-mail: 13151999695@163.com
林果园 男, (1975-), 博士, 副教授. 研究方向为云计算, 信息安全等.