

基于通道间相关性的图像重着色检测

陈诺¹, 祁树仁¹, 张玉书¹, 薛明富¹, 花忠云²

(1. 南京航空航天大学计算机科学与技术学院, 江苏 南京 211106; 2. 哈尔滨工业大学(深圳)计算机科学与技术学院, 广东 深圳 518055)

摘要: 图像重着色是一种新兴的图像编辑技术, 通过篡改像素值达到改变图像颜色风格的目的。随着社交网络和图像编辑技术的快速发展, 重着色图像已经严重阻碍了信息传达的真实性。然而, 专门为重着色而设计的工作少之又少, 现有的重着色检测方法在传统重着色场景下仍有很大提升空间, 在应对手工重着色图像时效果不佳。为此, 提出了一种基于通道间相关性的重着色图像检测方法, 该方法适用于重着色任务中的传统重着色和手工重着色场景。基于相机成像和重着色图像生成方式之间存在显著差异这一现象, 提出重着色操作或许会破坏自然图像的通道间相关性这一假设。通过数值分析说明, 通道间相关性差异可作为区分重着色图像和自然图像的重要鉴别度量。基于上述先验知识, 所提方法通过提取差分图像的一阶微分残差的通道共生矩阵, 获得图像的通道间相关性特征集。此外, 根据实际情况, 假设了3种检测场景, 包括训练-测试数据之间匹配、不匹配以及手工重着色场景。实验结果表明, 所提方法能够准确识别重着色图像, 在假设的3种场景下均优于现有方法, 取得了较高的检测精度。除此之外, 所提方法对训练数据量的依赖性较小, 在训练数据有限的情况下, 能实现相当精确的预测结果。

关键词: 图像重着色; 篡改检测; 通道间相关性; 图像取证

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2022057

Image recoloring detection based on inter-channel correlation

CHEN Nuo¹, QI Shuren¹, ZHANG Yushu¹, XUE Mingfu¹, HUA Zhongyun²

1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

2. College of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China

Abstract: Image recoloring is an emerging editing technique that can change the color style of an image by modifying pixel values. With the rapid proliferation of social networks and image editing techniques, recolored images have seriously hampered the authenticity of the communicated information. However, there are few works specifically designed for image recoloring. Existing recoloring detection methods still have much improvement

收稿日期: 2022-05-25; 修回日期: 2022-07-01

通信作者: 张玉书, yushu@nuaa.edu.cn

基金项目: 南京航空航天大学研究生科研与实践创新计划项目 (xcxjh20211606)

Foundation Item: Graduate Research and Practice Innovation Program Project of Nanjing University of Aeronautics and Astronautics (xcxjh20211606)

引用格式: 陈诺, 祁树仁, 张玉书, 等. 基于通道间相关性的图像重着色检测[J]. 网络与信息安全学报, 2022, 8(5): 167-178.

Citation Format: CHEN N, QI S R, ZHANG Y S, et al. Image recoloring detection based on inter-channel correlation [J]. Chinese Journal of Network and Information Security, 2022, 8(5): 167-178.

space in conventional recoloring scenarios and are ineffective in dealing with hand-crafted recolored images. For this purpose, a recolored image detection method based on inter-channel correlation was proposed for conventional recoloring and hand-crafted recoloring scenarios. Based on the phenomenon that there were significant disparities between camera imaging and recolored image generation methods, the hypothesis that recoloring operations might destroy the inter-channel correlation of natural images was proposed. The numerical analysis demonstrated that the inter-channel correlation disparities can be used as an important discriminative metric to distinguish between recolored images and natural images. Based on such new prior knowledge, the proposed method obtained the inter-channel correlation feature set of the image. The feature set was extracted from the channel co-occurrence matrix of the first-order differential residuals of the differential image. In addition, three detection scenarios were assumed based on practical situations, including scenarios with matching and mismatching between training-testing data, and scenario with hand-crafted recoloring. Experimental results show that the proposed method can accurately identify recolored images and outperforms existing methods in all three hypothetical scenarios, achieving state-of-the-art detection accuracy. In addition, the proposed method is less dependent on the amount of training data and can achieve fairly accurate prediction results with limited training data.

Keywords: image recoloring, forgery detection, inter-channel correlation, image forensics

0 引言

随着智能手机和移动互联网的普及，每一个人不仅是信息的接收者还是信息的生产者与传播者。图像相比文字符号，具有更强的还原现实和展现生活的能力，被广泛分享到微信、微博等社交媒体上。然而，随着 AdobePhotoshop 等图像编辑软件的迅速发展，任何人都能以极低的成本对图像进行篡改，并且产生的内容从视觉上难以分辨真假^[1]。除了对图像的内容进行拼接、复制-粘贴、修补、删除等篡改操作^[2]，最新的图像编辑技术（如图像重着色），还可以改变图像的风格。

图像重着色不改变原始图像内容，而通过篡改像素值的方式来改变图像的主题或风格^[3]。自 2001 年以来，各种图像重着色方法被提出。根据其运行机制大致可分为两类：传统的重着色^[3-17]和基于深度学习的重着色^[18-27]。大多数传统的重着色方法具有高度可定制化和低成本的优点，但需要根据特定图像和重着色区域动态地调整参数。基于深度学习的重着色方法可以根据图像的高级语义信息学习适当的特征表示，但由于目标分割算法的局限性，在小目标上的重着色效果显著下降，且所需的样本数目较大、运算时间都较长。

虽然大多数人仅出于娱乐目的对图像进行重着色，但由于颜色对于视觉系统至关重要，这种类型的篡改严重阻碍了信息传达的真实性。图像

重着色示例如图 1 所示，此种肉眼难辨真伪的重着色图像不仅可能误导个人及公众的意见形成，甚至可能干扰与安全密切相关的人工视觉系统的正常工作。

因此，迫切需要数字图像内容取证技术来鉴定数字图像的真实性和完整性。近年来，学术界在数字图像盲取证领域取得了显著进展^[28]。盲取证，又被称为被动取证，不依赖数字水印^[29]等任何形式的先验知识，仅依赖所分析的数字图像，并试图揭示在数字相机成像阶段或图像编辑阶段产生的特定伪像。例如，Yerushalmy 等^[30]利用透镜和传感器之间的相互作用所产生的像差进行伪造检测。一些研究利用图像域^[31-32]或小波域^[33]的统计工具分析伪造图像的局部噪声。由于复制-粘贴篡改在图像中有相同区域，Fridrich 等^[34]在 2003 年提出针对复制-粘贴篡改的检测工作。

总体来说，现有的取证方法在检测传统篡改方面取得了良好的性能。但是将这些方法直接应用于重着色场景的效果并不令人满意，因为它们没有考虑到重着色过程产生的特定伪像。虽然改变颜色是图像处理中最常见的操作之一，但专门为重着色图像而设计的取证工作仍然很少。Yan 等^[35]尝试从图像的通道间相关性和光照一致性的角度来应对传统的重着色。尽管他们的方法表现出良好的泛化性能，但没有充分发掘出通道间相关性的潜力，仍有很大提升空间。紧随其后，

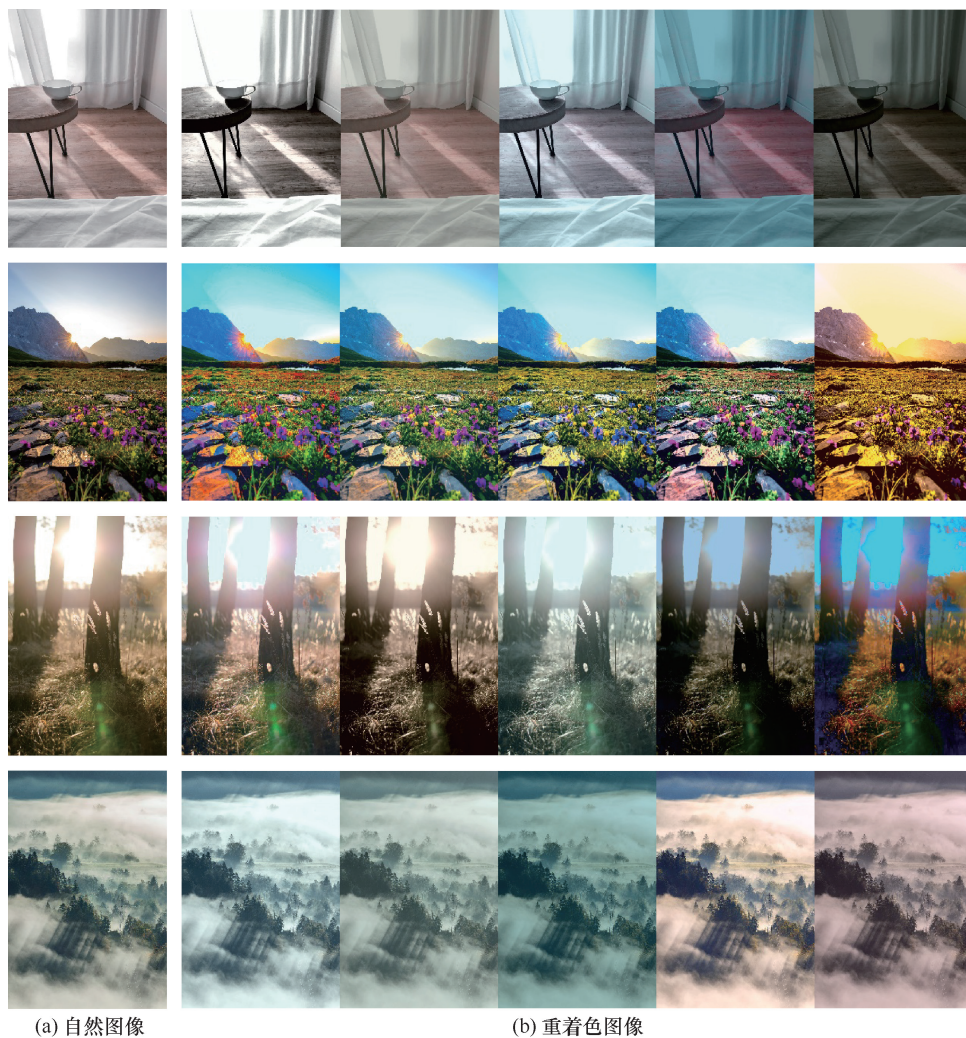


图 1 图像重着色示例
Figure 1 Examples of image recoloring

Zhang 等^[36]提出了一种基于空间相关性的重着色检测方法，该方法显示出了对传统和深度学习重着色的通用检测能力，但在应对手工制作的重着色场景时效果不佳。

为了缓解来自传统重着色的挑战及提高在手工重着色场景下的泛化能力，本文充分挖掘图像的通道间相关性，提出了一种有效且具有可解释性的检测方法来区分自然图像(NI, natural image)和重着色图像(RI, recolored image)。

本文的工作始于如下假设：由于相机成像和图像重着色方式之间存在内在差异，重着色操作或许会破坏 R 、 G 、 B 通道之间的相关性。随后的数值分析实验验证了这一假设。基于上述先验知识，为了更好地表示通道间相关性，首先使用高通滤波抑制通道间图像中与伪像无关的内容，然

后提取其中的通道间相关性特征。

本文的主要贡献总结如下。

- 1) 提出了一种基于通道间相关性的重着色图像检测方法，该方法能够有效缓解来自传统重着色面临的问题，提高在手工重着色场景下的泛化能力。
- 2) 通过理论分析得出，重着色操作或许会破坏自然图像的通道间相关性。此外，数值实验还验证了通道间相关性是一种具有区分性的特征。
- 3) 提出了一种有效的通道间相关性特征集用于重着色检测。

1 相关工作

1.1 传统的重着色方法

根据篡改机制，传统的重着色方法可大致分为：基于示例的重着色^[3-9]、基于编辑传播的重着色

色^[10-14]、基于调色板的重着色^[15-17]。

基于示例的重着色方法通过将参考图像的颜色风格迁移到目标图像,实现重着色。早期方法^[3]通过分析图像的全局颜色分布,将参考图像的特征应用到目标图像中,这类方法属于全局颜色迁移。此外,局部颜色迁移方法^[5]首先识别目标图像和参考图像之间的区域相关性,然后在对应的区域之间迁移颜色分布。

基于编辑传播的方法要求用户首先使用所需的颜色在图像的特定区域涂鸦;然后,算法将这些编辑自动传播到邻近区域中的相似像素。此类方法具有高度可定制的优点,缺点是需要动态地调整参数到特定的图像和区域。

基于调色板的方法根据输入图像的内容生成一个合适的调色板。用户只需要替换调色板中的颜色,图像的颜色样式就会自动改变。大多数方法使用聚类算法来定制特定图像的调色板,因此不适合大批量生成图像。

1.2 重着色检测方法

迄今为止,很少有专门区分自然图像和重着色图像的工作。Yan 等^[35]尝试为重着色图像设计检测算法。他们发现,从 RGB 图像中得到的通道间图像^[37]和光照图像^[38]可能对伪造检测^[39]有效。因此,利用上述特征,设计了一个基于深度神经网络^[40]的重着色图像检测模型,该模型由 3 个特征提取模块、一个特征融合模块和一个分类模块组成。首先,将原始 RGB 图像、通道间图像和光照图像分别输入一个特征提取分支(深度学习神经网络的前 3 个卷积阶段);在前一阶段提取与伪造相关的伪像特征后,采用连接层来融合这些特征;最后,在分类阶段获得图像经过重着色的概率。

从检测结果来看,该方法可以在一定程度上识别重着色图像。然而,在制作精良的手工数据集上,他们的方法仅能获得 65% 的准确率。随后, Zhang 等^[36]提出了一个基于空间相关性的通用检测方法,通过理论和数值分析发现重着色操作或许会破坏像素之间的空间相关性。基于这样的先验知识,设计了一个识别重着色图像的特征集,该特征集使用共生矩阵统计每个颜色分量上相邻像素之间的空间相关性;然后,将该特征集送入卷积神经网络中学习高级表征;最后,输出一张图片经过重着色篡改的概率。虽然该方法显示出了在传统和基于深度学习的重着色场景下的通用检测能力,但由于没有充分考虑到传统重着色的特定伪像,检测效果仍有很大的提升空间。

2 本文方法

本文提出了一个简单有效的方法用于检测重着色图像,总体框架如图 2 所示。该方法分为两个阶段:特征提取阶段和分类阶段。在特征提取阶段,首先,为了过滤与特定伪像无关的内容信息,依次获得原始图像的差分图像和残差图像;然后,提取残差图像的通道间相关性并组成特征集,记作通道共生矩阵。在分类阶段,使用支持向量机(SVM, support vector machine)训练分类器,该分类器根据提取的相关性特征判断图像是否经过重着色篡改。

2.1 通道间相关性的分析

相机成像与重着色图像的生成方式之间具有很大的区别。在自然图像中,由于成像过程中的彩色滤波阵列插值算法,不同颜色通道间存在高度相关性。然而,对于重着色图像,由于重着色操作中

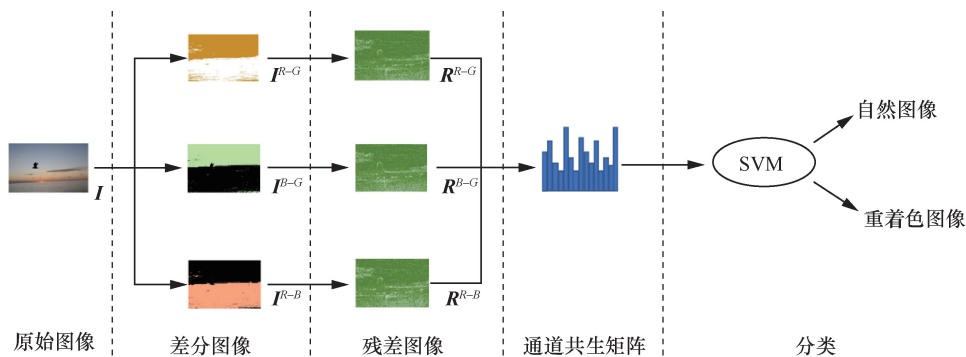


图 2 所提方法的总体框架

Figure 2 The overall framework of the proposed method

R 、 G 、 B 这 3 个颜色通道的数值变化往往是不同的, 所以 3 个通道间的相关性往往会遭受一定程度的破坏。因此, 与自然图像不同的是, 重着色图像中不同颜色通道之间并不是高度相关的。在接下来的分析中, 将展示一些实验证据来支持这一假设。

虽然重着色图像与自然图像在视觉上难以区分, 但重着色过程可能会改变自然图像的统计特征, 如颜色通道间的相关性。因此, 本文使用一种度量方法来展现自然图像和重着色图像的 R 、 G 、 B 颜色分量间的通道相关性差异。

对于数据集中的任意图像 I , 利用式(1)计算 I 的颜色通道间的相关系数:

$$r^{c_1, c_2} = \frac{\sum_{j,k} (I_{j,k}^{c_1} - \bar{I}^{c_1})(I_{j,k}^{c_2} - \bar{I}^{c_2})}{\sqrt{\sum_{j,k} (I_{j,k}^{c_1} - \bar{I}^{c_1})^2 \sum_{j,k} (I_{j,k}^{c_2} - \bar{I}^{c_2})^2}} \quad (1)$$

其中, w 和 h 分别表示图像的宽度和高度, j 和 k 为图像中元素的位置索引, $\bar{I}^c$ 是 I^c 的均值, r^{c_1, c_2} 表示图像 I 的颜色分量 c_1 和 c_2 之间的相关系数, 其值越大, 表示 c_1 和 c_2 之间的相关性越强, $c_1, c_2 \in \{R, G, B\}$ 。

对于一个重着色数据集, 首先计算其中每幅图像 I 任意颜色分量 c_1 和 c_2 之间的相关系数 r^{c_1, c_2} ; 然后根据该系数构造一个直方图, 记为 $\mathcal{H}_{\text{rec}}^{c_1, c_2}$ 。同理, 对于一组自然图像, 构造了直方图 $\mathcal{H}_{\text{nat}}^{c_1, c_2}$ 。然后, 使用卡方距离来衡量直方图 $\mathcal{H}_{\text{rec}}^{c_1, c_2}$ 和 $\mathcal{H}_{\text{nat}}^{c_1, c_2}$ 之间的相似性, 如式(2)所示。

$$d_{\chi^2} = \frac{1}{2} \sum_x \frac{(\mathcal{H}_{\text{rec}}^{c_1, c_2}(x) - \mathcal{H}_{\text{nat}}^{c_1, c_2}(x))^2}{\mathcal{H}_{\text{rec}}^{c_1, c_2}(x) + \mathcal{H}_{\text{nat}}^{c_1, c_2}(x)} \quad (2)$$

其中, x 为直方图的收集箱的索引, d_{χ^2} 可以作为反映自然图像和重着色图像之间统计差异的鉴别度量, 其值越高, 表示自然图像和重着色图像之间的统计差异越大。

为了评估鉴别度量 d_{χ^2} 的有效性, 使用训练集 D_1 进行一些分析实验。首先, 使用式(1)分别计算 D_1 中所有图像的 B 与 G 、 B 与 R 、 G 与 R 颜色通道之间的相关系数 r^{c_1, c_2} ; 然后, 分别获得相应的直方图 $\mathcal{H}_{\text{rec}}$ 和 $\mathcal{H}_{\text{nat}}$ (如图 3 所示), 并计算 d_{χ^2} 。

$\mathcal{H}_{\text{rec}}$ 和 $\mathcal{H}_{\text{nat}}$ 的非重叠区域反映了重着色图像和

自然图像之间的差异。所有子图的鉴别度量 d_{χ^2} 均超过 0.06。这些结果表明, 重着色图像和自然图像在通道间相关性方面存在统计学差异, 若有效地提取这些统计特征将有利于重着色图像的检测工作。

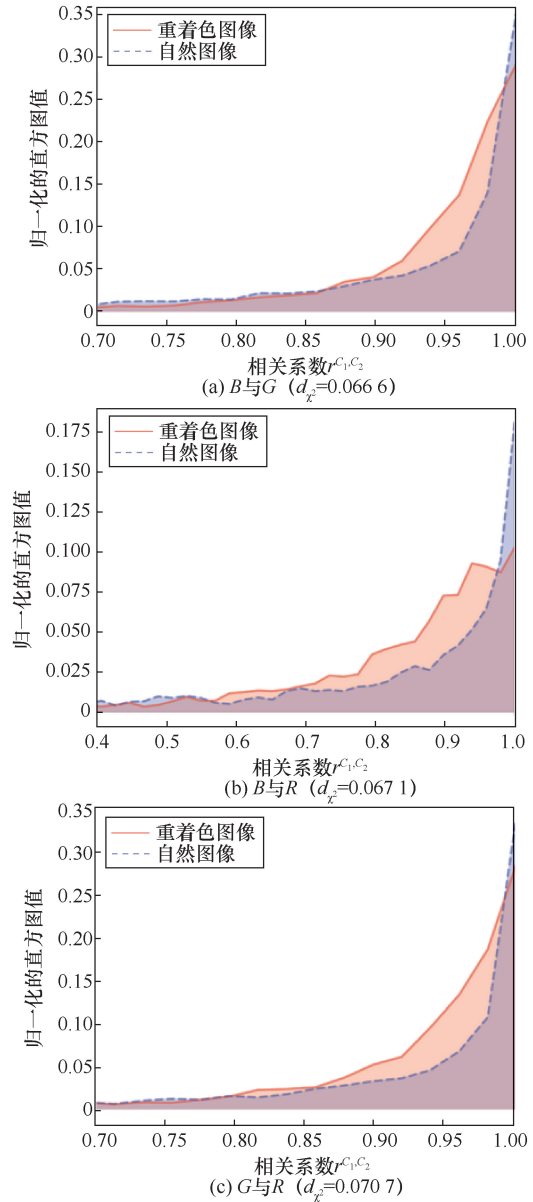


图3 重着色图像和自然图像的通道间相关性直方图的差异
Figure 3 Differences in inter-channel correlation histograms of recolored images and natural images

2.2 提取通道间相关性特征

步骤1 计算差分图像

基于上述分析, 充分挖掘差分图像的通道间相关性特征。Gunturk 等^[41]提出, 自然图像的不同颜色通道的高频分量间具有很强的相关性。图

像的一个颜色通道可表示为:

$$I^c = I_l^c + I_h^c \quad (3)$$

其中, $c \in \{R, G, B\}$, l 和 h 分别表示图像颜色通道的高频和低频分量。差分图像可表示为:

$$I^{c_1} - I^{c_2} = I_l^{c_1} + I_h^{c_1} - I_l^{c_2} - I_h^{c_2} \quad (4)$$

其中, c_1 和 c_2 分别表示不同的颜色通道。

对于自然图像, 由于高频分量的相似性, $I_h^{c_1} \approx I_h^{c_2}$ 。所以, 差分图像可表示为:

$$I^{c_1} - I^{c_2} \approx I_l^{c_1} - I_l^{c_2} \quad (5)$$

因此, 在自然图像的差分图像中, 相应的高频分量被过滤, 只保留低频分量。

对于重着色图像, 由关于通道间相关性的分析结果可知, 不同颜色通道间并不是高度相关的。因此, 差分图像不能使用式(5)表示, 即不能过滤高频分量。相比自然图像, 重着色图像的差分图像包含更多的高频分量。在本项工作中, 计算每幅图像的 3 个差分图像, 即 $R-G$ 、 $B-G$ 和 $R-B$, 并将它们沿通道维度堆叠。

步骤 2 计算差分图像的一阶微分残差

由上述分析可知, 自然图像的差分图像中, 只包含低频分量; 而重着色图像中除了包含低频分量外, 还包含高频分量。两类图像中均包含的低频分量可能对二者之间差异的分析产生负面影响。因此, 首先采用高通滤波抑制差分图像的低频分量, 然后研究残差图像之间的差异更具合理性。

本文使用一阶微分算子获得残差图像:

$$R_{j,k}^c = DI_{j,k}^c - DI_{j,k+1}^c \quad (6)$$

其中, $c \in \{R-G, B-G, R-B\}$, DI^c 是差分图像的第 c 个分量, R^c 是相应的残差图像。式(6)仅为水平方向的残差图像, 竖直方向的残差图像计算公式同理可得。

步骤 3 截断残差图像

通过步骤 2, 分别获得了差分图像在竖直方向和水平方向的残差图像。然而残差图像中有太多不同的元素值, 如果直接对原始残差数据提取通道间相关性特征, 会使获得的通道共生矩阵维数较大, 增加了运算的复杂度。因此, 在计算通道相关性之前, 还需要对残差图像进

行预处理。

为了减少残差图像中不同元素值的数量, 使用式(7)对残差图像 R^c 进行截断:

$$\hat{R}_{j,k}^c = \begin{cases} \tau, & R_{j,k}^c \geq \tau \\ R_{j,k}^c, & -\tau < R_{j,k}^c < \tau \\ -\tau, & R_{j,k}^c \leq -\tau \end{cases} \quad (7)$$

其中, $\tau > 0$ 为截断阈值。截断后, 得到的残差图像 $\hat{R}^c$ 中只包含 $[-\tau, \tau]$ 的整数值。

步骤 4 提取通道间相关性特征

受灰度共生矩阵的启发, 使用式(8)计算残差图像的通道间相关性, 记作通道共生矩阵:

$$C(\theta_1, \theta_2, \theta_3) = \frac{1}{n} \sum_{j,k} \mathbb{I} \begin{pmatrix} \hat{R}_{j,k}^{R-G} = \theta_1 \\ \hat{R}_{j,k}^{B-G} = \theta_2 \\ \hat{R}_{j,k}^{R-B} = \theta_3 \end{pmatrix} \quad (8)$$

其中, $\mathbb{I}()$ 为指示函数, n 是归一化因子, θ_1 、 θ_2 、 θ_3 是通道共生矩阵的索引。 $\hat{R}^c$ 的每个通道共生矩阵的维数为 $(2\tau + 1)^3$ 。

由于通道共生矩阵是对称的, 所以可以将 $C(\theta_1, \theta_2, \theta_3)$ 和 $C(\theta_3, \theta_2, \theta_1)$ 合并来减少特征维度。组合后, $\hat{R}^c$ 的每个通道共生矩阵的维度大幅缩减为 $((2\tau + 1)^3 + (2\tau + 1)^2) / 2$ 维。特征维数的降低是图像的自然对称性引起的, 因此不会显著降低检测性能。事实上, 由此产生的相对低维的特征将加速分类器的训练。

在本文的实验中, 设置 $\tau = 2$, 分别计算每幅图像水平方向残差和竖直方向残差的通道共生矩阵, 然后将这两个通道共生矩阵逐元素求和, 作为最终的通道相关性特征, 维度为 75。在获得通道相关性特征后, 使用支持向量机来训练并预测图像是否经过重着色篡改。

3 实验

3.1 训练集和评价指标

为了方便后文的描述, 本文提出的方法记作 ICC-Method (method based on inter-channel correlation)。实验使用 Zhang 等^[36]生成的重着色数据集作为训练集来训练 ICC-Method。使用的数据集由两个子集组成, 记作 D_1 。第一个子集包括 15 000 幅

自然图像,是从 ImageNet 验证数据集^[42]中随机抽样获得的,使用重着色方法^[3]生成相应的重着色图像。第二个子集具有与第一个子集相同的容量,自然图像也是从 ImageNet 验证数据集^[42]中随机抽样获得的。对应的重着色图像是使用重着色方法^[7]生成的。训练集 D_1 包含 1 000 个类别的大量室内室外的真实场景,能够很好地模拟真实的重着色场景,示例如图 4 所示。本文所有数据集中的自然图像和重着色图像都不重复。实验中使用 3 个数据集 (D_2 、 D_3 、 D_4) 用于测试,如表 1 所示。



图 4 训练集 D_1 的示例

Figure 4 Examples selected from the training set D_1

表 1 实验中使用的数据集汇总
Table 1 Summary of datasets used in the experiments

数据集	用途	重着色方法	重着色图像数目	自然图像来源	自然图像数目
D_1	训练	文献[3]	18 000	ImageNet ^[42]	18 000
		文献[7]	18 000	ImageNet ^[42]	18 000
D_2	测试	文献[3]	60	ImageNet ^[42]	60
		文献[7]	60	ImageNet ^[42]	60
D_3	测试	文献[3,6-9,12,17]	100	下载自网络	100
D_4	测试	手工制作	80	—	—

为了评估 ICC-Method 在重着色检测中的性能,本文使用准确率和受试者工作特征 (ROC, receiver operating characteristic) 曲线下的面积 (AUC, area under the curve) 两个指标对实验结果进行分析。

准确率指的是正确预测样本数在测试样本总数中的比例,其计算如下:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (9)$$

其中, TP 为将自然图像预测为自然图像, TN 为将重着色图像预测为重着色图像, FP 为将重着色图像预测为自然图像, FN 为将自然图像预测为重着色图像。

ROC 曲线的横坐标为假阳性率 (FPR, false positive ratio), 纵坐标是真阳性率 (TPR, true positive ratio)。二者的计算如下:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (10)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

3.2 参数选择

在评估 ICC-Method 与各种重着色方法的性能之前,先通过实验手工调整超参数,找到一个最优的超参数组合。本文实验从数据集 D_1 中随机选择 3 000 幅重着色图像及其对应的自然图像进行参数选择。本文实验使用了带有高斯核函数的 LIBSVM^[43], 其中有两个参数 c 和 g , 分别表示正则化因子和核系数。参数调优通过使用十折交叉验证的网格搜索实现。图 5 显示了 ICC-Method 在不同的 c 和 g 设置下的准确率。由图 5 得, ICC-Method 的最优参数分别是 $c = 32$, $g = 32$ 。图 6 展示了 ICC-Method 在 $c = 32$, $g = 32$ 时的交叉验证结果。由图 6 可得, ICC-Method 取得了良好的性能,平均准确率为 89.61%, 并且每一折的结果都非常近似。

3.3 性能评估

获得最优参数后,使用训练集 D_1 训练 ICC-Method, 并进行全面的性能评估。实际上,由于重着色方法具有多种类型,不同方法产生的重着色图像之间可能存在分布偏差。因此,为了更全面地评估 ICC-Method 的有效性,根据测试数据的来源,分成 3 个场景进行检测: 训练集与测试集匹配的场景、训练集与测试集不匹配的场景、手工重着色场景。3 个场景下使用的测试集分别为 D_2 、 D_3 、 D_4 。

场景 1 训练集与测试集匹配

在本场景中,本文评估了训练数据和测试数据匹配的情况下, ICC-Method 在测试集 D_2 上的性能。训练数据和测试数据匹配,指的是训练集和测试集中的重着色图像是由相同的重着色方法使用相同来

c 值	$g=1/64$	$g=1/32$	$g=1/16$	$g=1/8$	$g=1/4$	$g=1/2$	$g=1$	$g=2$	$g=4$	$g=8$	$g=16$	$g=32$	$g=64$
$c=1/64$	63.06%	63.03%	63.20%	63.61%	65.03%	65.73%	65.31%	65.51%	65.98%	66.43%	67.26%	68.58%	70.55%
$c=1/32$	63.06%	63.03%	63.20%	64.98%	65.80%	65.71%	65.25%	65.83%	66.90%	68.13%	69.78%	72.41%	74.33%
$c=1/16$	63.06%	63.03%	64.95%	65.70%	65.85%	65.68%	66.05%	66.85%	68.86%	70.75%	73.70%	76.23%	77.56%
$c=1/8$	63.06%	65.01%	65.66%	66.01%	65.85%	66.06%	66.73%	68.81%	71.83%	74.71%	77.30%	79.00%	79.58%
$c=1/4$	64.93%	65.56%	66.21%	66.30%	66.08%	66.66%	68.35%	72.33%	75.91%	78.18%	79.91%	81.60%	82.56%
$c=1/2$	65.55%	66.08%	66.41%	66.31%	67.00%	68.30%	71.28%	76.25%	78.45%	80.78%	82.68%	83.81%	84.71%
$c=1$	66.10%	66.41%	66.75%	67.16%	68.40%	70.96%	75.38%	78.55%	81.16%	83.31%	84.51%	85.85%	86.65%
$c=2$	66.33%	66.90%	67.35%	68.50%	70.55%	73.83%	78.28%	80.86%	83.15%	85.13%	86.43%	87.26%	87.88%
$c=4$	66.98%	67.31%	68.76%	70.50%	73.15%	77.20%	80.65%	82.63%	85.20%	86.71%	87.75%	88.30%	88.58%
$c=8$	67.50%	68.76%	70.35%	72.80%	75.90%	80.25%	82.53%	84.66%	86.63%	87.81%	88.60%	88.93%	88.88%
$c=16$	68.75%	70.43%	72.45%	75.33%	78.78%	82.21%	84.28%	86.43%	87.86%	88.85%	89.36%	89.21%	88.93%
$c=32$	70.45%	72.41%	75.08%	78.05%	82.00%	83.66%	85.76%	87.46%	88.90%	89.35%	89.45%	89.61%	89.21%
$c=64$	72.31%	74.85%	77.56%	81.00%	83.53%	85.03%	87.28%	88.60%	89.31%	89.53%	89.53%	89.07%	89.48%

图 5 ICC-Method 在不同 SVM 参数设置下的准确率
Figure 5 Accuracy of ICC-Method for different SVM parameter settings

源的自然图像生成的。测试集 D_2 由 Zhang 等^[36]使用重着色方法^[3,7]生成, 包含两个子集, 每个子集包含重着色图像及相应的自然图像共 120 幅, 使用由 Zhang 等^[36]提出的与训练集 D_1 相同的重着色方法生成。如表 2 所示, 准确率和 AUC 分别达到了 93.75% 和 96.88%。这些结果表明, 当分类器经过相应的数据训练后, 可以准确地识别重着色图像。

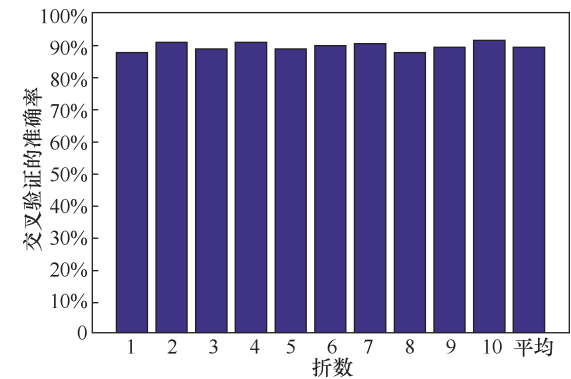


图 6 ICC-Method 的十折交叉验证的准确率
Figure 6 Accuracy of ICC-Method with 10-fold cross validation

表 2 ICC-Method 在不同测试集上的结果
Table 2 Results of ICC-Method on different testing sets

测试集	准确率	AUC
D_2	93.75%	96.88%
D_3	90.50%	94.68%
D_4	85.00%	—

场景 2 训练集与测试集不匹配

实际上, 研究人员可能并不完全知道图像来源或图像经何种重着色方法篡改过。在这种情况下, 训练数据和测试数据之间存在不匹配, 包括两种情况: 第一种情况, 训练数据和测试数据中的图片来源不同; 第二种情况, 训练数据和测试数据经过不同的重着色方法生成。

本场景使用 Yan 等^[35]生成的数据集作为本节实验的测试集, 记作 D_3 。测试集 D_3 中包括 100 幅下载自网络的真实图像, 然后使用各种重着色方法^[3,6-9,12,17]生成相应的重着色图像。为了测试检测方法应对重着色操作的泛化性, 基于编辑传播和调色板的重着色方法也被用于生成重着色图像。图 7 显示了 D_3 中一些自然图像和相应的重着色例子。



(a) 自然图像 (b) 重着色图像

图 7 从 D_3 中随机选择的图像示例
Figure 7 Example images randomly selected from D_3

为了更好地说明 ICC-Method 的优越性, 本文将其与目前的 4 种方法进行了比较。

1) RecDeNet (recolored detection network): Yan 等^[35]专门为传统重着色篡改而设计的检测方法。该方法使用原始的 RGB 图像及其派生的通道间图像和光照图像来检测图像是否经过传统重着色篡改。

2) SC-Method (method based on spatial correlation): Zhang 等^[36]提出的应对传统和深度重着色的通用检测方法。该方法通过卷积神经网络从空间相关特征集中学习特征表示。

3) RF-Method (method based on rich features): Goljan 等^[44]为图像隐写分析而设计的方法。该方法在图像取证方面取得了良好的性能。该方法首先提取 SCRMQ1 特征, 然后训练一个 SVM 进行分类。

4) DCC-Method (method based on disparities in color components): Li 等^[45]提出的应对生成对抗网络 (GAN, generative adversarial network) 生成图像的识别问题。该方法假设 GAN 生成的图像与自然图像在颜色分量上存在差异, 因此可用于检测图像重着色篡改。

表 3 展示了所有方法在测试集 D_3 上的比较结果, 图 8 为相应的 ROC 曲线。请注意, 由于 RecDeNet 和 RF-Method 的结果是直接引用 Yan 等^[35]的数据 (二者在测试集 D_4 上的结果也是直接引用), 因此缺少相应的 ROC 曲线。由于 RF-Method 和 DCC-Method 不是专门为图像重着色任务而设计的, 因此效果最差。SC-Method 是通用图像重着色检测方法, 效果相比于 RecDeNet 和 ICC-Method 仍有显著差距。ICC-Method 取得了最佳的检测结果, 说明相比 RecDeNet, 该方法更加充分地挖掘了图像的通道间相关性, 所提出的通道共生矩阵特征能够有效地区分出自然图像和重着色图像。

表 3 在测试集 D_3 上的比较结果
Table 3 Comparison results on the testing set D_3

方法	准确率	AUC
RecDeNet	83.50%	90.75%
SC-Method	77.00%	84.35%
RF-Method	69.00%	—
DCC-Method	51.50%	50.18%
ICC-Method	90.50%	94.68%

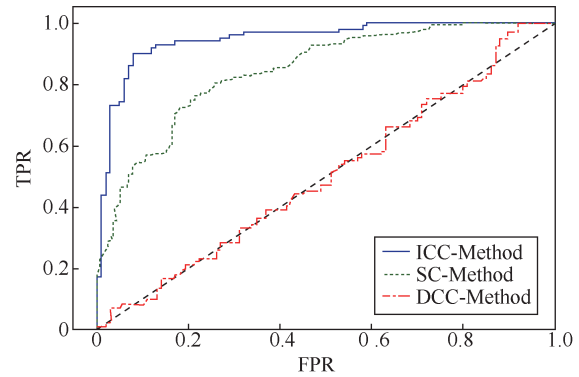


图 8 在测试集 D_3 上的 ROC 曲线
Figure 8 ROC curves on testing set D_3

场景 3 手工重着色场景

实际上, 重着色图像不仅可以通过各种重着色技术来生成, 还可以通过各种图像编辑软件, 如 AdobePhotoshop 和 GIMP 来生成。由于人们具有的先验知识, 手工制作的重着色图像比重着色技术自动生成的图像留下的伪像更少 (如图 9 所示), 这是一项更具挑战性的任务。本场景下使用 Yan 等^[35]公开的手工数据集, 记作 D_4 。 D_4 包含 80 幅手工制作的重着色图像, 通过手机应用制作而成, 或者从 Photoshop 教程等网站上下下载获得。5 种方法在测试集 D_4 上的比较结果如表 4 所示。由于 D_4 仅包含重着色图像, 因此 AUC 无法计算得到。除 DCC-Method 外, 其余方法在 D_4 上的准确率相比于在 D_3 上都有所下降, 其中 RF-Method 下降幅度最大, 为 31.5%, 而 ICC-Method 下降幅度最小, 仅 5.5%。即使手工重着色图像的真实性很高, 检测难度很大, ICC-Method 仍取得了最高的准确率, 比 DCC-Method 高 2.5%。除此之外, DCC-Method 在 D_3 上的准确率最低, 而 ICC-Method 的准确率达到 90.5%。上述结果表明, 通道间相关性更能代表来自不同来源的重着色图像的共同特征, 所提出的特征编码方法能充分发挥通道间相关性的潜力。



图 9 从 D_4 中随机选择的图像示例
Figure 9 Example images randomly selected from D_4

表 4 在测试集 D_4 上的比较结果
Table 4 Comparison results on the testing set D_4

方法	准确率
RecDeNet	65.00%
SC-Method	70.00%
RF-Method	37.50%
DCC-Method	82.50%
ICC-Method	85.00%

3.4 训练集数目对性能的影响

为了验证所提出的通道间相关性特征在少量训练样本上的优势, 本文进行了一个额外的实验。通过使用不同数量的训练图像对, 观察 ICC-Method 分别在测试集 D_2 、 D_3 、 D_4 上的准确率的变化情况。如图 10 所示, 仅需 150 对训练图像, ICC-Method 就能在测试集 D_2 和 D_3 上获得超过 80% 的准确率。即使是在难度很大的手工重着色测试集 D_4 上, ICC-Method 也仅需要 1 500 对训练图像就能获得 81.2% 的准确率。这些结果表明 ICC-Method 对训练数据量的依赖性较小。即使在重着色图像数量很有限的现实场景下, ICC-Method 也能实现相当精确的预测结果。

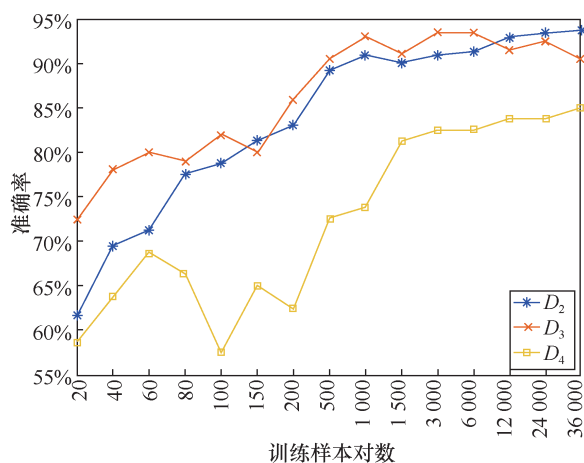


图 10 不同数量的训练样本对的准确率
Figure 10 The accuracy of different numbers of training sample pairs

3.5 失败案例

虽然 ICC-Method 在多种传统重着色方法生成的测试集 D_2 、 D_3 及手工重着色测试集 D_4 上取得了相当精确的检测精度, 但通道相关性特征并不总是有效。Zhang 等^[36]使用深度重着色方法^[19,24-26]生成了一个重着色测试集, 其中包含

480 幅重着色图像及其对应的自然图像。ICC-Method 在其上的准确率仅为 69.46%, 比 SC-Method 低 15.75%。一些失败案例如图 11 所示。假负例表示 ICC-Method 将该自然图像分类为重着色图像, 而假正例表示未被检测出的重着色图像。在深度重着色场景下, ICC-Method 的准确率显著下降, 可能是由于深度重着色方法采用的自动编码器和生成对抗网络等深度学习工具能够很好地学习自然图像的通道间相关性, 所提出的通道间相关性特征集难以捕获自然图像与深度重着色图像之间的通道间相关性差异。在未来的工作中, 将通过探索更能代表不同来源的重着色图像的共同特征来解决这个问题。



图 11 失败案例
Figure 11 Failure cases

3.6 实际场景应用

重着色技术是把“双刃剑”, 在给人们带来便利的同时也产生了负面的影响。在特定的背景和应用场合下, 肉眼难辨真伪的重着色图像不仅可能误导个人及公众的意见形成, 甚至可能干扰司法, 对社会造成很大的影响。如据美联社 5 月 26 日报道, 有网络用户在社交网络推特上称美国广播公司在报道得克萨斯州校园枪击案时, 对枪手的照片进行了包括重着色在内的编辑, “使他看起来更像白人”。虽然随后被证实为谣言, 但截至该推特被平台处理时, 该谣言已获得 2 500 多个转发和近 4 000 个赞。

为此, 本文从 COCO 数据集^[46]中挑选了 60 张人物肖像照片, 其中包含不同的肤色、性别、年龄段、场景。随后使用图像编辑软件“美图秀秀”

对每张图片进行了提亮肤色、添加滤镜等重着色操作。虽然重着色图像肉眼难辨真伪, 但 ICC-Method 仍能获得 90.83% 的准确率。

上述结果表明本文提出的基于通道间相关性的检测方法能够为现实世界的重着色相关纠纷提供技术支持。

4 结束语

本文提出了一种基于通道间相关性的重着色图像检测方法, 适用于重着色任务中的传统重着色和手工重着色场景。基于相机成像和重着色图像生成方式之间的显著差异, 假设重着色操作或许会破坏自然图像的通道间相关性。随后的数值分析表明, 虽然自然图像和重着色图像在视觉上难以分辨, 但通道间相关性差异可作为区分二者的重要鉴别度量。基于上述先验知识, 本文提取差分图像的一阶微分残差的通道共生矩阵特征, 并且在 3 种不同假设情况下, 评估了 ICC-Method 的有效性。实验结果表明, 即使是训练数据与测试数据不匹配和手工重着色这两种具有挑战性的场景, ICC-Method 也能取得较高的检测准确率, 所提出的通道间相关性特征具有良好的泛化能力。同时, ICC-Method 对训练数据量的依赖性较小。希望本文关于通道间相关性的发现能为重着色检测任务及相关的图像取证任务提供有用的见解。在未来, 将探索更具有区别性的特征和更有效的特征编码方法来提高检测性能。

参考文献:

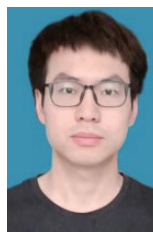
- [1] NIGHTINGALE S J, WADE K A, WATSON D G. Can people identify original and manipulated photos of real-world scenes[J]. *Cognitive Research: Principles and Implications*, 2017, 2(1): 30.
- [2] VERDOLIVA L. Media forensics and DeepFakes: an overview[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2020, 14(5): 910-932.
- [3] REINHARD E, ADHIKMIN M, GOOCH B, et al. Color transfer between images[J]. *IEEE Computer Graphics and Applications*, 2001, 21(5): 34-41.
- [4] PITIE F, KOKARAM A C, DAHYOT R. N-dimensional probability density function transfer and its application to color transfer[C]//*Proceedings of Tenth IEEE International Conference on Computer Vision (ICCV'05)*. 2005: 1434-1439.
- [5] TAI Y W, JIA J, TANG C K. Local color transfer via probabilistic segmentation by expectation-maximization[C]//*Proceedings of 2005 IEEE Conference on Computer Vision and Pattern Recognition*. 2005: 747-754.
- [6] BEIGPOUR S, VAN DE WEIJER J. Object recoloring based on intrinsic image estimation[C]//*Proceedings of 2011 International Conference on Computer Vision*. 2011: 327-334.
- [7] PITIE F, KOKARAM A C, DAHYOT R. Automated colour grading using colour distribution transfer[J]. *Computer Vision and Image Understanding*, 2007, 107(1/2): 123-137.
- [8] PITIE F, KOKARAM A. The linear monge-kantorovitch linear colour mapping for example-based colour transfer[C]//*European Conference on Visual Media Production*. 2007: 1-9.
- [9] GROGAN M, PRASAD M, DAHYOT R. L2 registration for colour transfer[C]//*Proceedings of 2015 23rd European Signal Processing Conference (EUSIPCO)*. 2015: 1-5.
- [10] LEVIN A, LISCHINSKI D, WEISS Y. Colorization using optimization[M]//*ACM Siggraph*. 2004: 689-694.
- [11] QU Y, WONG T T, HENG P A. Manga colorization[J]. *ACM Transactions on Graphics (TOG)*, 2006, 25(3): 1214-1220.
- [12] AN X B, PELLACINI F. AppProp: all-pairs appearance-space edit propagation[C]//*Proceedings of ACM SIGGRAPH 2008*. 2008: 1-9.
- [13] XU K, LI Y, JU T, et al. Efficient affinity-based edit propagation using KD tree[J]. *ACM Transactions on Graphics (TOG)*, 2009, 28(5): 1-6.
- [14] CHEN X W, ZOU D Q, LI J W, et al. Sparse dictionary learning for edit propagation of high-resolution images[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 2854-2861.
- [15] O'DONOVAN P, AGARWALA A, HERTZMANN A. Color compatibility from large datasets[M]//*ACM Siggraph*. 2011: 1-12.
- [16] LIN S, RITCHIE D, FISHER M, et al. Probabilistic color-by-numbers: Suggesting pattern colorizations using factor graphs[J]. *ACM Transactions on Graphics (TOG)*, 2013, 32(4): 1-12.
- [17] CHANG H, FRIED O, LIU Y, et al. Palette-based photo recoloring[J]. *ACM Transactions on Graphics (TOG)*, 2015, 34(4): 1-11.
- [18] TSAI Y H, SHEN X H, LIN Z, et al. Sky is not the limit: semantic-aware sky replacement[J]. *ACM Transactions on Graphics*, 2016, 35(4): 1-11.
- [19] LEE J Y, SON H, LEE G, et al. Deep color transfer using histogram analogy[J]. *The Visual Computer*, 2020, 36(10): 2129-2143.
- [20] HE M, LIAO J, CHEN D, et al. Progressive color transfer with dense semantic correspondences[J]. *ACM Transactions on Graphics (TOG)*, 2019, 38(2): 1-18.
- [21] ENDO Y, IIZUKA S, KANAMORI Y, et al. DeepProp: extracting deep features from a single image for edit propagation[J]. *Computer Graphics Forum*, 2016, 35(2): 189-201.
- [22] ZHANG R, ZHU J Y, ISOLA P, et al. Real-time user-guided image colorization with learned deep priors[J]. *arXiv preprint arXiv:1705.02999*, 2017.
- [23] CHO J, YUN S, LEE K, et al. PaletteNet: image recolorization with given color palette[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2017: 1058-1066.
- [24] AFIFI M, PRICE B L, COHEN S, et al. Image recoloring based on object color distributions[C]//*Eurographics (Short Papers)*. 2019: 33-36.
- [25] LUAN F, PARIS S, SHECHTMAN E, et al. Deep photo style transfer[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 4990-4998.
- [26] LI Y J, LIU M Y, LI X T, et al. A closed-form solution to photorealistic image stylization[C]//*Proceedings of the European Conference*

- on Computer Vision. 2018: 453-468.
- [27] YOO J, UH Y, CHUN S, et al. Photorealistic style transfer via wavelet transforms[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 9035-9044.
- [28] BIRAJDAR G K, MANKAR V H. Digital image forgery detection using passive techniques: a survey[J]. Digital Investigation, 2013, 10(3): 226-245.
- [29] 吴金海, 林福宗. 基于数字水印的图像认证技术[J]. 计算机学报, 2004, 27(9): 1153-1161.
- WU J H, LIN F Z. Image authentication based on digital watermarking[J]. Chinese Journal of Computers, 2004, 27(9): 1153-1161.
- [30] YERUSHALMY I, HEL-OR H. Digital image forgery detection based on lens and sensor aberration[J]. International Journal of Computer Vision, 2011, 92(1): 71-91.
- [31] POPESCU A C, FARID H. Statistical tools for digital forensics[C]//International Workshop on Information Hiding. 2004: 128-147.
- [32] LYU S W, PAN X Y, ZHANG X. Exposing region splicing forgeries with blind local noise estimation[J]. International Journal of Computer Vision, 2014, 110(2): 202-221.
- [33] MAHDIAN B, SAIC S. Using noise inconsistencies for blind image forensics[J]. Image and Vision Computing, 2009, 27(10): 1497-1503.
- [34] FRIDRICH A J, SOUKAL B D, LUKÁŠ A J. Detection of copy-move forgery in digital images[C]//Digital Forensic Research Workshop. 2003.
- [35] YAN Y Y, REN W Q, CAO X C. Recolored image detection via a deep discriminative model[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(1): 5-17.
- [36] ZHANG Y S, CHEN N, QI S R, et al. Detecting recolored image by spatial correlation[J]. arXiv preprint arXiv:2204.10973, 2022.
- [37] HO J S, AU O C, ZHOU J T, et al. Inter-channel demosaicking traces for digital image forensics[C]//Proceedings of 2010 IEEE International Conference on Multimedia and Expo. 2010: 1475-1480.
- [38] VAN DE WEIJER J, GEVERS T, GIJSENIJ A. Edge-based color constancy[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2007, 16(9): 2207-2214.
- [39] CARVALHO T, FARIA F A, PEDRINI H, et al. Illuminant-based transformed spaces for image forensics[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(4): 720-733.
- [40] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [41] GUNTURK B K, ALTUNBASAK Y, MERSEREAU R M. Color plane interpolation using alternating projections[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2002, 11(9): 997-1013.
- [42] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [43] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 27.
- [44] GOLJAN M, FRIDRICH J, COGRANNE R. Rich model for Steganalysis of color images[C]//Proceedings of 2014 IEEE

International Workshop on Information Forensics and Security. 2014: 185-190.

- [45] LI H D, LI B, TAN S Q, et al. Identification of deep network generated images using disparities in color components[J]. Signal Processing, 2020, 174: 107616.
- [46] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context[C]//European Conference on Computer Vision. 2014: 740-755.

[作者简介]



陈诺 (1996-), 男, 安徽六安人, 南京航空航天大学硕士生, 主要研究方向为数字图像内容取证。



祁树仁 (1994-), 男, 辽宁朝阳人, 南京航空航天大学博士生, 主要研究方向为视觉表征、稳健模式识别和媒体内容安全。



张玉书 (1987-), 男, 甘肃庆阳人, 南京航空航天大学教授、博士生导师, 主要研究方向为多媒体安全与人工智能、区块链与物联网安全、云计算与大数据安全。



薛明富 (1986-), 男, 江苏南京人, 南京航空航天大学副教授, 主要研究方向为人工智能安全、硬件安全、硬件木马检测和深度学习模型的版权保护。



花忠云 (1989-), 男, 湖南湘西州人, 哈尔滨工业大学 (深圳) 副教授、博士生导师, 主要研究方向为混沌理论及应用、多媒体安全、信息隐藏和图像处理。