

基于差分卷积和通道注意力的 JPEG 彩色图像隐写分析

王方馨¹ 丁云瑶¹ 雷善中¹ 王爱鑫¹

WANG Fangxin DING Yunyao LEI Shanzhong WANG Aixin

摘要

目前大多数隐写分析方法均针对灰度图像进行设计,无法有效检测彩色图像。为解决这一问题,文章提出了基于差分卷积和通道注意力的 JPEG 彩色图像隐写分析方法,首先利用高通滤波器获取各个通道的隐写信息,其次设计差分卷积模块,引入中心差分卷积提取纹理复杂区域的特征,并引入通道注意力机制 SENet 获取各个通道特征的重要程度。实验结果表明,所提方法针对隐写算法 J-UNIWARD 和 UERD 的检测精度均有明显提升,相比 SRNet 最大提升了 1.55%,并且通过实验证明所提方法的模型复杂度更低。

关键词

JPEG 隐写分析; 彩色图像; 高通滤波器; 差分卷积; 通道注意力机制

doi: 10.3969/j.issn.1672-9528.2025.01.016

0 引言

图像隐写术旨在尽可能保证图像视觉效果和统计特性的情况下,巧妙地将秘密信息嵌入载体之中^[1]。而图像隐写分析作为一种对抗手段,要求专业人士对图像进行深入分析,以发现潜在的隐写痕迹。随着隐写术的不断进步,图像隐写分析也变得越来越复杂和困难,图像隐写分析领域面临着巨大挑战。

图像隐写分析分为传统的隐写分析和基于深度学习的隐写分析两大类。其中,传统的隐写分析依赖于专业人员的经验和对图像特征的深入理解来构建特征,费时费力且精度较低;基于深度学习的隐写分析通过构建端到端的网络实现隐写分析,优势是能够通过自动学习并提取深层次的图像特征,无须借助手工设计的特征。最为常见的方法^[2-11]是利用多层卷积层来提取图像残差和全连接层进行特征分类的结构,并取得了比传统隐写分析方法更好的检测性能。

然而,目前大多数基于深度学习的隐写分析网络设计仅针对灰度图像,且针对灰度图像的隐写分析网络在处理彩色图像时,无法准确识别出隐写前后各个颜色通道之间的统计差异,导致检测性能并不尽如人意。从目前的研究进展看,针对彩色图像的隐写分析的工作相对较少,其检测性能未达到较好效果。JPEG 图像作为一种广泛使用的图像格式,因其在网络传输和存储中的普遍性,成为隐写分析研究的重要对象。

为了有效提升 JPEG 彩色图像隐写分析方法

的检测精度,本文提出了一种基于差分卷积和通道注意力的 JPEG 彩色图像隐写分析方法。首先利用高通滤波器获取各个通道的隐写信息,其次利用中心差分卷积层代替普通卷积层构建差分卷积模块提取隐写特征,并嵌入通道注意力机制 SENet 自适应地学习通道特征的重要程度。实验结果表明,本文方法相对于 SRNet,在 JPEG 彩色图像隐写分析的检测精度上有一定的提升。

1 本文方法

本节介绍了所提出的 JPEG 彩色图像隐写分析网络,该网络由 3 个核心部分构成,下面小节中详细阐述了这 3 个部分。

1.1 网络结构

本文提出的网络总体结构如图 1 所示,所提出的网络包含了三个阶段,分别是预处理阶段、特征提取阶段和分类阶段。

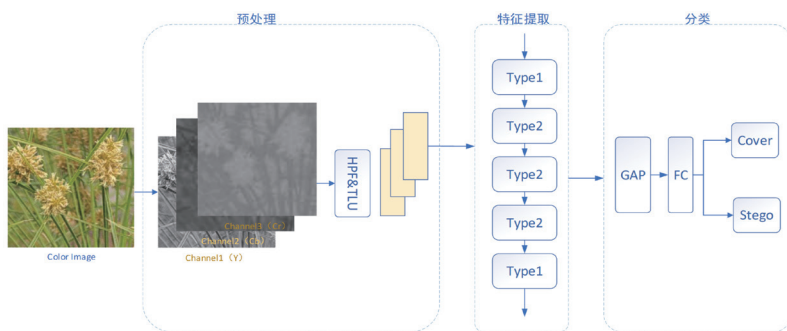


图 1 网络结构

预处理阶段首先进行通道分离,其次利用高通滤波器获取各个通道的高频信息并串联以保留更多的隐写嵌入;特征提取阶段采用了 2 个卷积模块(如图 2(a))和 3 个

1. 西藏民族大学信息工程学院 陕西咸阳 712082

[基金项目] 西藏民族大学研究生科研创新与实践项目 (Y2024055 Y2024056)

差分卷积模块（如图 2（b））提取更深层次的隐写特征，组合方式为 1 个卷积模块和 3 个差分卷积模块，最后再添加 1 个卷积模块；分类阶段则利用池化层和全连接层对获取的隐写特征进行分类，判断输入图像是否包含隐写信息。

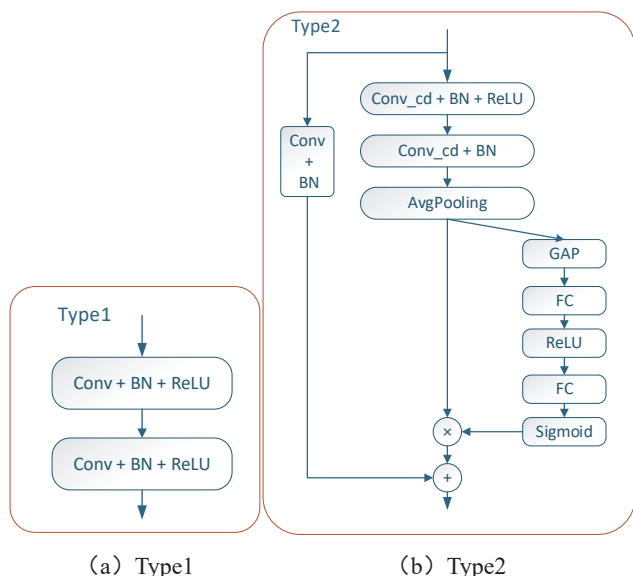


图 2 特征提取阶段

1.2 预处理阶段

预处理阶段主要用来提取图像的残差信息，提高隐写信息和图像内容的比例。隐写术通常需要对图像进行微小的修改，这些修改会破坏图像中相邻像素之间原有的相关性，从而影响后续的特征提取过程。通过预处理获取这种被破坏的相关性，利于后续特征提取。因此，引入高通滤波器不仅有助于减少由于噪声引起的误差累积，还能促进网络前期的稳定性和收敛性。在预处理阶段，首先需要将输入的 JPEG 彩色图像分成 Y、Cb 和 Cr 三个通道，然后对这三个通道分别使用 30 个固定的 SRM 高通滤波核进行卷积，得到相应的残差信息，并对获取的残差信息使用截断阈值为 4 的截断线性操作（TLU）减小残差的动态范围，最后将各个通道的残差信息串联在一起形成一个包含 90 维特征图的矩阵。

该阶段巧妙地运用 SRM 高通滤波器来处理图像数据，旨在发掘出图像中被忽略但具有重要意义的高通残差信息，有利于后期隐写特征的提取。同时为了保留各个通道内在的相关性，将各个通道的残差信息串联在一起，能够有效保留各个通道嵌入的隐写信息，有利于后续的特征提取。

1.3 差分卷积模块

为了加强对纹理复杂区域的特征提取，本文利用中心差分卷积代替普通卷积应用于特征提取模块，有助于模型提取更深层次的隐写特征。相比于普通卷积，中心差分卷积能够

更好地描述细粒度地纹理信息，而自适应 JPEG 隐写算法更倾向于将隐写信息嵌入纹理复杂的区域，如图 3 所示。

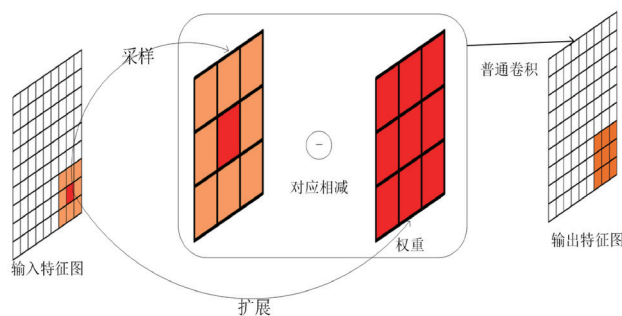


图 3 中心差分卷积计算过程

中心差分卷积的采样操作与普通卷积一致，但聚合阶段更侧重于采样值的中心向梯度，具体公式为：

$$y(p_0) = \theta \cdot \sum_{p_n \in R} \omega(p_n) \cdot (x(p_0 + p_n) - x(p_0)) + (1 - \theta) \cdot \sum_{p_n \in R} \omega(p_n) \cdot x(p_0 + p_n) \quad (1)$$

式中： x 、 y 分别为中心差分卷积的输入与输出特征图； p_0 表示输入特征图的当前位置； p_n 表示 p_0 的邻域； R 表示卷积运算的不同方向向量； θ ($\theta \in [0, 1]$) 为超参数，本文所设置的 θ 为 0.5。

对于整个差分卷积模块，其结构如图 2 中 Type2 所示，首先利用两个中心差分卷积层进行局部特征提取，为了提高模型的稳定性，在第一个卷积层后加入 BN 层和 ReLU 激活函数，在第二个卷积层后加入 BN 层，并增加短连接层防止隐写信息经过多层卷积后消失，同时嵌入了通道注意力机制 SENet 使模型自主地区分通道特征的重要程度。

1.4 通道注意力机制 SENet

卷积神经网络在挖掘特征时，会为每个通道特征设定一个默认的权重，这意味着其重要性是相同的。且目标参数受不同通道特征的影响程度不同，导致模型无法有效分配贡献度，影响隐写分析的检测精度。

因此本文引入了通道注意力机制 SENet，通过其内置的自适应算法来衡量各个通道特征的重要性程度，确保网络能够更加关注那些最具代表的关键信息，忽略那些相对次要或冗余的特征，实现隐写分析检测精度的提升。

SENet 主要包括挤压（squeeze）和激励（excitation）两个操作。挤压操作通过全局平均池化将包含全局信息的 $W \times H \times C$ 的特征图压缩为 $1 \times 1 \times C$ 的特征向量，该特征向量具有全局感受野，其计算式为：

$$z = F_{sq}(X_c) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (2)$$

式中： H 、 W 为特征图大小； $X_c \in \mathbf{R}^{H \times W \times c}$ 为输入特征图； c 为通道数； z 为挤压操作生成的权重。

激励操作的作用是提取特征通道间的相关性，采用两个全连接层实现。第一个全连接层起降维作用，减少网络参数并增强模型泛化能力，第二个全连接层可以使其原始维度恢复。其计算过程可表示为：

$$S_c = F_{ex}(Z, W) = \sigma[W_2 \delta(W_1 Z)] \quad (3)$$

式中： δ 为Relu函数； σ 为Sigmoid函数； $S_c \in \mathbf{R}^{1 \times 1 \times c}$ 为生成的注意力权重。

最后将原始特征与学习到的各通道权值相乘，其计算公式为：

$$c = F_{scale}(X_c, S_c) = S_c \otimes X_c \quad (4)$$

式中： $\otimes$ 代表逐元素相乘； X_c 为输出特征图。

2 实验结果及分析

2.1 数据集

本文实验采用隐写分析领域常用彩色图像数据集 ALASKA II，该数据集各个图像的大小为 $256 \text{ px} \times 256 \text{ px}$ 。本次实验从该数据集随机选择 20 000 张图像作为实验样本，选取两种 JPEG 彩色图像隐写算法 J-UNIWARD 和 UERD，首先将载体图像压缩为质量因子为 75 px 的 JPEG 图像，并选取 3 种隐写嵌入率获得相应的载秘图像。最终，划分训练集、验证集和测试集的比例为 14:1:5。

2.2 实验设置

本次实验操作系统为 Windows11，深度学习框架为 PyTorch，优化器为 SGD 优化器，初始学习率是 0.02，经过两次衰减，140 个 epoch 之后，下降到 0.002，190 个 epoch 之后，下降到 0.000 02。在训练过程中，批量大小设定为 25。

2.3 对比试验

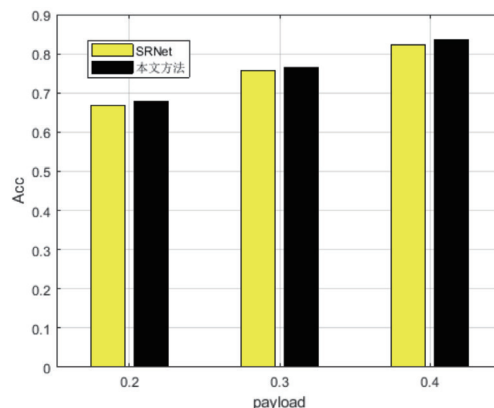
为了进一步验证本文方法的有效性，在本节中，将所提出的 JPEG 彩色图像隐写分析模型与现有典型 JPEG 隐写分析方法进行了性能对比。主要是经过修改后的灰度图像隐写分析网络 SRNet，为了公平比较，对 SRNet 做了细微的修改，在第 1 层将 $H \times W$ 核修改为 $3 \times H \times W$ 核，其中， H 和 W 分别是卷积核的高和宽，3 为彩色图像的通道数。实验过程中，使用了两种 JPEG 隐写算法 J-UNIWARD 和 UERD，所采用的嵌入率为 0.4 bpnzac、0.3 bpnzac 和 0.2 bpnzac，实验结果如表 1 所示。通过表 1 可以看出，在不同嵌入率的情况下，本文方法的检测性能均优于 SRNet。当隐写算法为 UERD、嵌入率为 0.4 bpnzac 时，本文方法检测准确率最高，达到 88.84%，相比 SRNet 提升了 1.24%；当隐写算法为 J-UNIWARD、嵌入率为 0.2 bpnzac 时，本文方法

检测效果最差，检测准确率为 67.85%，相比 SRNet 提升了 1.09%。

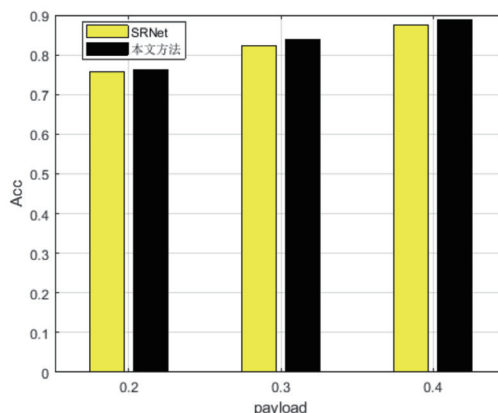
表 1 在质量因子 75 下针对 J-UNIWARD 和 UERD 算法，本文方法和 SRNet 的检测效果对比

隐写算法	模型	嵌入率 /bpnzac		
		0.2	0.3	0.4
J-UNIWARD	SRNet	0.667 5	0.756 5	0.821 5
	本文方法	0.678 4	0.764 3	0.837 0
UERD	SRNet	0.756 2	0.822 5	0.876 0
	本文方法	0.762 1	0.839 6	0.888 4

同时，通过图 4 可以看出，检测准确率和嵌入率之间存在正相关关系，即随着隐写操作的嵌入率不断提升，原始图像中所引入的隐写扰动也随之增加，隐写信息更易被检测到。



(a) 对 J-UNIWARD



(b) 对 UERD

图 4 质量因子 75 下对 J-UNIWARD 和 UERD，本文方法和 SRNet 检测效果对比

2.4 网络复杂度对比

在本节中，采用了两种不同的方法来评估本文提出的网络和 SRNet 的网络复杂度。具体而言，通过计算两个网络在

训练过程中所需的参数数量以及训练时间，进而对网络复杂度进行对比。表 2 详细列出了本文方法及其对比方法 SRNet 的参数数量和训练时间的对比数据。

表 2 本文方法和 SRNet 网络复杂度对比

	参数数量	训练时间 /h
SRNet	4 780 770	16.99
本文方法	651 362	12.36

从表 2 中可以明显看出，SRNet 相较于本文网络来说，其模型包含了更多的参数。这表明，本文提出的网络不仅具有更好的性能表现，还具有更低的模型复杂度、更短的训练时间和更高的计算效率。此外，由于参数数量的减少，网络更容易理解和解释，泛化能力更强，且在实际应用中更具可扩展性和灵活性。这些优势使得网络不仅在理论上具有吸引力，而且在实际应用中也更为实用，尤其是在计算资源受限或需要快速部署的场景中。

3 总结

为了提升 JPEG 彩色图像隐写分析的检测精度，本文提出了基于差分卷积和通道注意力的 JPEG 彩色图像隐写分析方法，该网络包括预处理阶段、特征提取阶段和分类阶段。其中，预处理阶段利用 30 个固定的 SRM 高通滤波核获取噪声残差信息并串联起来，提高了隐写信息和图像内容的比例；此外，特征提取阶段使用中心差分卷积代替普通卷积构建中心差分卷积模块，并嵌入了通道注意力机制 SENet 获取各个通道特征的重要程度，确保每个通道的特征都得到充分的利用。实验结果表明，在 JPEG 彩色图像隐写分析领域，本文方法与 SRNet 相比具有更好的检测性能和更小的模型复杂度。尽管目前 JPEG 彩色图像隐写分析已经取得了初步的成功，但在未来的研究中，将不断探索和创新，以期达到更高的性能标准。

参考文献：

[1] 陈君夫,付章杰,张卫明,等.基于深度学习的图像隐写分析综述[J].软件学报,2021,32(2): 551-578.

[2] XU G S, WU H Z, SHI Y Q. Structural design of convolutional neural networks for steganalysis[J]. IEEE signal processing letters, 2016, 23(5): 708-712.

[3] YE J, NI J Q, YANG Y. Deep learning hierarchical representations for image steganalysis[J]. IEEE transactions on information forensics and security,2017,12(11): 2545-2557.

[4] ZHANG R, ZHU F, LIU J Y, et al. Depth-wise separable convolutions and multi-level pooling for an efficient spatial

CNN-based steganalysis[J]. IEEE transactions on information forensics and security, 2019, 15: 1138-1150.

[5] YOU W K, ZHANG H, ZHAO X F. A siamese CNN for image steganalysis[J]. IEEE transactions on information forensics and security, 2020, 16: 291-306.

[6] ZENG J S, TAN S Q, LI B, et al. Pre-training via fitting deep neural network to rich-model features extraction procedure and Its effect on deep learning for steganalysis[J]. Electronic imaging, 2017(7): 44-49.

[7] XU G S. Deep convolutional neural network to detect J-UNWARD [C]// Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security.NewYork:ACM, 2017: 67-73.

[8] BOROUMAND M, CHEN M, FRIDRICH J. Deep residual network for steganalysis of digital images[J]. IEEE transactions on information forensics and security, 2018, 14(5): 1181-1193.

[9] WANG Z, CHEN M Z, YANG Y, et al. Joint multi-domain feature learning for image steganalysis based on CNN[J/OL]. EURASIP journal on image and video processing, 2020[2024-01-19].<https://jivp-eurasipjournals.springeropen.com/articles/10.1186/s13640-020-00513-7>.

[10] SU A T, ZHAO X F, HE X L. Arbitrary-sized JPEG steganalysis based on fully convolutional Network [C]// International Workshop on Digital Watermarking. Cham:Springer,2021: 197-211.

[11] ZENG J S, TAN S Q, LIU G Q, et al. WISERNet: wider separate-then-reunion network for steganalysis of color images[J]. IEEE transactions on information forensics and security, 2019, 14(10): 2735-2748.

【作者简介】

王方馨（1999—），通信作者（email: w1186297956@163.com），女，陕西咸阳人，硕士研究生，研究方向：隐写分析。

丁云瑶（2000—），女，山东青岛人，硕士研究生，研究方向：人脸活体检测。

雷善中（1998—），男，四川广安人，硕士研究生，研究方向：人脸伪造检测。

王爱鑫（2000—），女，甘肃平凉人，硕士研究生，研究方向：人脸活体检测。

（收稿日期：2024-09-27）