融合残差块和注意力机制的 JPEG 压缩反取证

唐贝贝[1],陈 磊[2],李若宇[2]
1. 安徽理工大学 人工智能学院,安徽 淮南 232001
2. 淮南师范学院 计算机学院,安徽 淮南 232001

摘 要:目的 针对 JPEG 压缩反取证方法在生成图像质量与反取证性能之间平衡不足的问题,设计了一种融合多级残差块和通道注意力机制的反取证模型 RBAM-JAF,以提高生成图像的质量,促使反取证性能和图像质量之间达到更好的平衡。**方法** 采用基于生成对抗网络的框架,包括生成器和鉴别器两部分,其中,生成器引入多级残差块和通道注意力机制,使模型具有更好的泛化性,增强图像特征表示的能力;添加特征融合模块,以充分利用所有卷积层的特征,提高生成图像的质量。**结果** 实验结果显示,与现有的 4 种反取证方法 M_1、M_2、M_3 和 M_4 相比,在 $Q_F=25$ 的情况下,PSNR 值分别增长了 8.52%、3.31%、1.52%和 0.07%,SSIM 值分别增长了 12.89%、2.46%、1.90%和 0.55%;在 $Q_F=50$ 的情况下,PSNR 值分别增长了 10.22%、2.21%、0.88%和 0.19%,SSIM 值分别增长了 9.71%、1.52%、0.64%和 0.21%;在 $Q_F=75$ 的情况下,PSNR 值分别增长了 18.26%、3.56%、3.80%和 2.96%,SSIM 值分别增长了 10.83%、1.58%、1.16%和 0.52%。此外,通过 4 种取证检测器的检测,$Q_F=25$、50 和 75 时,AUC 值均接近或低于 0.5。**结论** 实验结果显示:方法 M_5 与现有的反取证方法相比,提高了生成图像的视觉质量且能够有效地欺骗现有的取证检测器,使反取证性能和生成图像质量之间达到了更好的平衡。

关键词:JPEG 压缩;JPEG 反取证;生成对抗网络;残差块;通道注意力机制
中图分类号:TP391.41 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2025.0003.008

JPEG Compression Anti-forensics Integrating Residual Blocks and Attention Mechanism

TANG Beibei[1], CHEN Lei[2], LI Ruoyu[2]
1. School of Artificial Intelligence, Anhui University of Science and Technology, Anhui Huainan 232001, China
2. School of Computer Science, Huainan Normal University, Anhui Huainan 232001, China

Abstract: Objective To address the lack of balance between the quality of generated images and anti-forensics performance in JPEG compression anti-forensics methods, an anti-forensics model RBAM-JAF combining multi-level residual blocks and channel attention mechanism was designed. This model aimed to improve the quality of generated images and achieve a better balance between anti-forensics performance and image quality. **Methods** A framework based on generative adversarial networks (GANs) was employed, including a generator and a discriminator. The generator incorporated multi-level residual blocks and channel attention mechanisms to enhance the model's generalization capability and improve the representation of image features. Additionally, a feature fusion module was introduced to fully utilize

收稿日期:2023-10-23 **修回日期:**2023-12-06 **文章编号:**1672-058X(2025)03-0063-07
基金项目:安徽省高校科研重点项目(2022AH051582);认知智能全国重点实验室开放课题资助(COGOS-2023HE02);淮南市 50 科技之星创新团队项目(623076).
作者简介:唐贝贝(1999—),女,安徽六安人,硕士研究生,从事计算机视觉研究.
通信作者:陈磊(1980—),男,安徽怀远人,教授,从事软件与数据科学研究. Email:leichen@hnnu.edu.cn.
引用格式:唐贝贝,陈磊,李若宇. 融合残差块和注意力机制的 JPEG 压缩反取证[J]. 重庆工商大学学报(自然科学版),2025,42(3):63-69.
TANG Beibei, CHEN Lei, LI Ruoyu. JPEG compression anti-forensics integrating residual blocks and attention mechanism[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(3): 63-69.

features from all convolutional layers, in order to enhance the quality of generated images. **Results** According to the experimental results, compared with four existing anti-forensic methods (M_1, M_2, M_3, and M_4), the proposed method showed significant improvements. At $Q_F = 25$, the PSNR values increased by 8.52%, 3.31%, 1.52%, and 0.07% respectively, and the SSIM values increased by 12.89%, 2.46%, 1.90%, and 0.55% respectively. At $Q_F = 50$, the PSNR values increased by 10.22%, 2.21%, 0.88%, and 0.19% respectively, and the SSIM values increased by 9.71%, 1.52%, 0.64%, and 0.21% respectively. At $Q_F = 75$, the PSNR values increased by 18.26%, 3.56%, 3.80%, and 2.96% respectively, and the SSIM values increased by 10.83%, 1.58%, 1.16%, and 0.52% respectively. Additionally, the AUC values of the four detectors for $Q_F = 25$、50, and 75 were close to or below 0.5. **Conclusion** Experimental results demonstrate that method M_5 improves the visual quality of generated images compared with existing methods while effectively deceiving forensic detectors, achieving a better balance between anti-forensic performance and the quality of the generated images.

Keywords: JPEG compression; JPEG anti-forensics; generative adversarial networks; residual blocks; channel attention mechanism

1 引 言

图像反取证是一种新开发的技术,旨在通过相应的后处理操作消除或模糊数字图像操作留下的痕迹,欺骗图像取证检测器,大大降低其取证性能。JPEG 压缩格式是如今使用最广泛的有损压缩格式之一,JPEG 压缩可能会引入阻塞伪影、振铃效应、模糊、DCT 频域和量化伪影等,降低了图像的视觉质量,由此提出了 JPEG 反取证技术[1]。该技术能够通过一定的处理去除或隐藏 JPEG 压缩留下的痕迹,恢复图像的视觉质量,从而欺骗现有的 JPEG 取证算法。通过对 JPEG 反取证的研究,可以发现目前取证技术的不足,并推动 JPEG 取证工具的进一步发展[2]。

JPEG 反取证方法主要是消除压缩痕迹的相关特征,从而达到欺骗取证算法的目的。该方法包括基于传统的反取证方法和基于深度学习的反取证方法。Stamm 等[3]开创了 JPEG 反取证工作,提出一种通过在 DCT 系数中添加适当噪声来消除量化伪影的方法;Stamm 等[4]在先前工作的基础之上提出将中值滤波和高斯白噪声引入到 JPEG 压缩图像中来掩盖阻塞伪影;Fan 等[5]开发了一种基于全变差(Total Variation,TV)的四步方法,包括 JPEG 去块痕迹处理、感知 DCT 直方图平滑、第二轮基于 TV 的去块痕迹处理和去校准 4 个步骤,以去除空间域和 DCT 域中的 JPEG 压缩伪影,并在取证的不可检测性和重建图像的质量之间达到良好的平衡。以上都是基于传统的 JPEG 反取证方法。随着深度学习的快速发展,基于生成对抗网络(Generative Adversarial Network,GAN)的 JPEG 反取证方法也逐渐被提出。Luo 等[1]首次提出基于生成对抗网络的 JPEG 压缩反取证研究,将高频率信息作为鉴别器的输入,以捕获生成图像和未压缩图像之间的统计差异;Wu 等[6]提出将面向 JPEG 压缩反取证问题看作图像到图像的转换问题,并使用基于高频 DCT 系数的损失函数来重建图像的高频分量,提高重建图像的视觉质量,同时引入校准损失,以减轻生成图像与其校准版本之间的高频子带方差差距;Wu 等[7]在上述研究的基础之上提出了 JRA-GAN,改进了高频 DCT 系数损失函数,并设计了一个合理的高频滤波器,以提高生成图像的质量[8]。

综上所述,现有的 JPEG 压缩反取证方法能够删除 JPEG 压缩留下的痕迹来欺骗现有的 JPEG 取证算法。但是,大多数现有的 JPEG 反取证方法仅追求反取证性能,而导致生成图像质量不佳。因此,如何在不牺牲图像视觉质量的情况下欺骗现有的取证检测器,仍然是一个具有挑战性的问题。针对上述问题,设计了 RBAM-JAF 模型,该框架由生成器(Generator,G)和鉴别器(Discriminator,D)构成。在生成器中,使用多级残差块(Residual Block,RB)[9],使所得到的模型具有更好的泛化性,可以实现更复杂的网络结构。其次,在多级残差块的末端加入通道注意力机制(Channel Attention Module,CA)[10-11],通过对特征通道之间的相互依赖关系,自适应地重新缩放每个通道特征,使得网络可以专注于更重要的通道信息。在 JPEG 反取证方法生成重建图像时,能够更好地保持细节、纹理等视觉质量,使得生成的图像更难被现有的取证检测器所发现。此外,随着网络深度的增加,残差特征难以被充分利用,在该框架中引入特征融合模块,促使其能够充分利用每个卷积层的特征,增强模型利用图像特征信息的能

力,从而提高生成图像质量。实验结果表明:RBAM-JAF 在峰值信噪比(Peak Signal-to-Noise Ratio,PSNR)和结构相似性(Structural Similarity,SSIM)[12]指标评估方面优于现有的 JPEG 反取证方法,且 AUC 值均接近或小于 0.5,在生成图像视觉质量和反取证性能方面,二者达到了更好的平衡。

其余部分安排如下:第二节介绍 JPEG 反取证方法,第三节介绍损失函数,第四节介绍仿真实验与结果分析,最后一节进行总结和展望。

2　JPEG 反取证方法

针对上述反取证问题,在 RBAM-JAF 框架中,将 JPEG 压缩图像生成重建图像,该框架由一个生成器 G 和一个鉴别器 D 组成。G 接收 JPEG 压缩图像并消除 JPEG 压缩留下的痕迹,生成重建图像,D 判别输入图像是原始未压缩图像还是由 G 生成的图像,具体框架由图 1 所示。

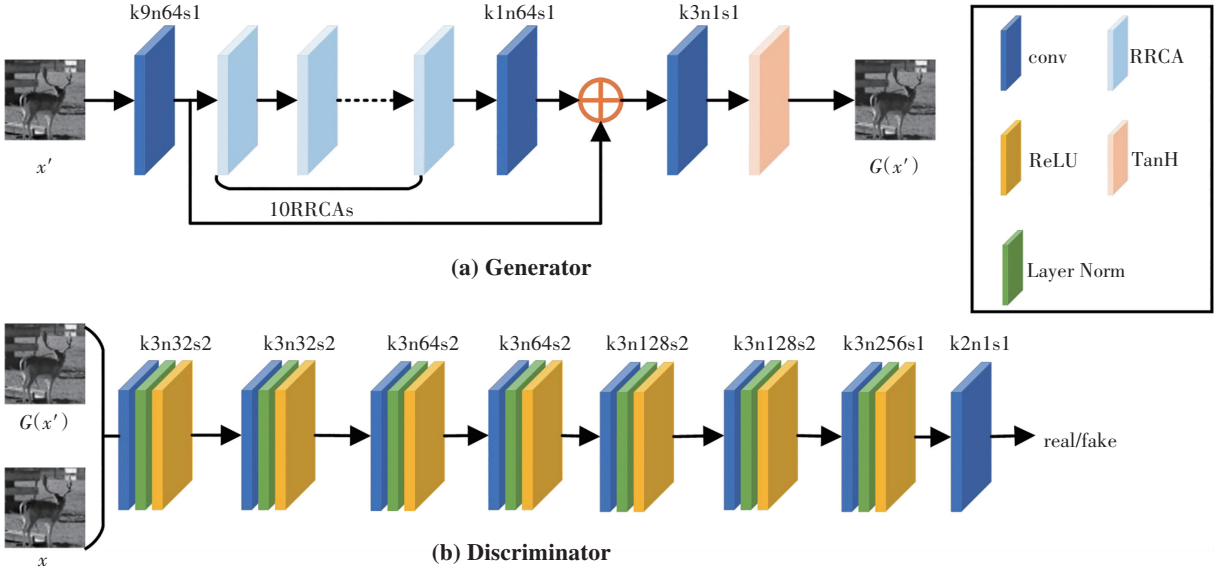

图 1　RBAM-JAF 网络模型结构

Fig. 1　Structure of RBAM-JAF network model

2.1　生成器

如上所述,G 用于生成 JPEG 反取证重建图像,为了生成具有良好视觉质量并与原始图像具有相似统计特征的重建图像,同时保证输出图像的大小与 G 输入图像的大小相同,采用全卷积神经网络。

G 的网络结构如图 1(a)所示。首先,输入一组 512×512 JPEG 压缩灰度图像到 Generator 中,通过第一层卷积层提取输入图像的低级特征,输入的像素值被缩放为[-1,1];然后,使用 10 个 RRCAs 有效地学习高频信息,重建提取的特征,再使用一个具有 64 个特征图和 3×3 个内核的卷积层进行全局特征提取,同时该层输出的特征图与第一层的输出求和进行全局残差学习,并传输到内核为 3×3 的卷积层中;最后,使用双曲正切函数 Tanh 作为激活函数将生成的图像像素值缩放到[-1,1]的范围内。在上述过程中,特征图的大小在每个卷积层中保持不变,步幅设置为 1。

RRCA 结构如图 2 所示。每个 RRCA 由 20 个 RBCAs 和一个具有 64 个特征图及 3×3 个内核的卷积层组成。第(m-1)个 RRCA 的输出作为第 m 个 RRCA 的输入,通过 20 个 RBCAs 提取特征传输到一个具有 64 个特征图和 3×3 个内核的卷积层进行特征学习;然后将第(m-1)个 RRCA 的输出与该层的输出结果相加,融合浅层信息和深层信息,减少信息的丢失,增强生成的图像信息,学习的结果输出作为第(m+1)个 RRCA 的输入。

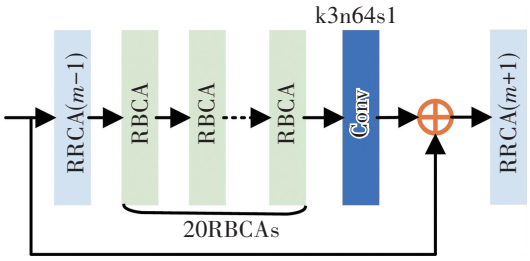

图 2　RRCA 模型结构

Fig. 2　RRCA model structure

RBCA 结构如图 3 所示,在多级残差块的基础之上,添加了通道注意力模块,由于通道之间的相互依赖,自适应地调整残差特征,增强图像特征表示的能力,使得 RBAM-JAF 网络专注于学习更有用的通道信息,在生成

反取证图像时更好地保持图像的细节、纹理等视觉质量,能够有效地欺骗现有的取证检测器。第($n-1$)个 RBCA 的输出为第 n 个 RBCA 的输入,经过 3 块卷积-ReLU-卷积的残差块操作,将输出的结果传入到通道注意力机制模块调整残差特征,且每个残差块的输出都输入到 64 个特征图和 1×1 卷积层中进行融合,使每个卷积层的特征能够被充分的利用,为生成重建图像提供更多的信息,提高生成图像的视觉质量。最后,将第($m-1$)个 RBCA 的输出与该层的输出相结合,作为第($n+1$)个 RBCA 的输入,其中所有的卷积都是 64 个特征图和 3×3 个内核。

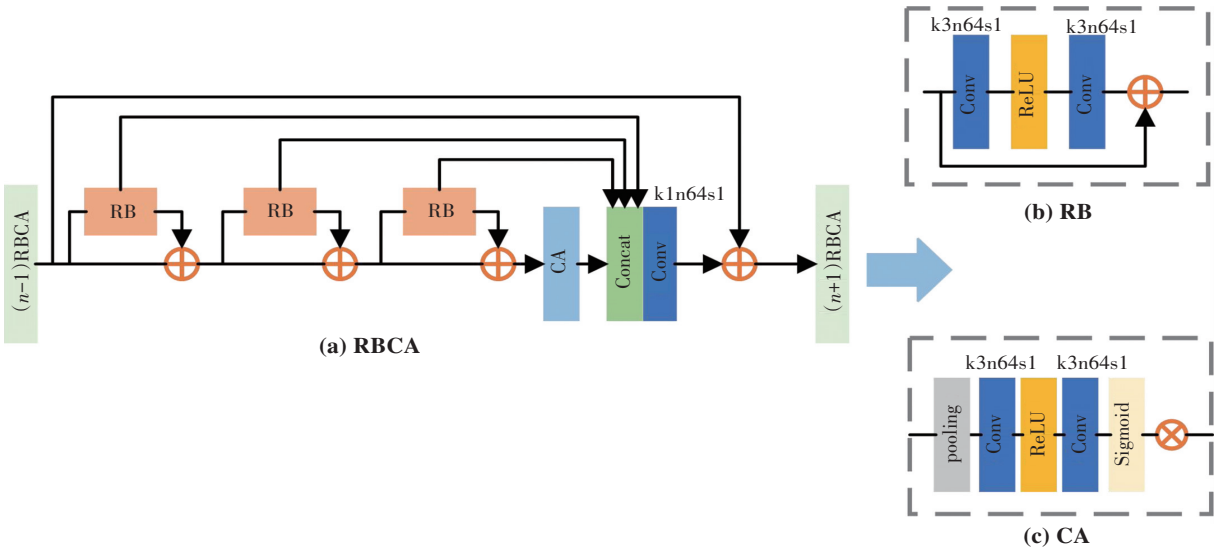

图 3 RBCA 模型结构

Fig. 3 RBCA model structure

2.2 鉴别器

Discriminator 结构如图 1(b)所示。首先,D 接收生成器生成的图像或原始未压缩图像作为输入,减去输入图像的像素值,将图像缩放到[-1,1]范围内;其次,利用一系列卷积层来学习更高级别的典型特征,每个卷积层都具有 3×3 个内核,每层后面都跟一个归一化层和一个 ReLU 激活函数,每两个卷积层滤波器数量加倍,最后一层卷积层具有 2×2 个内核;最后,D 输出一个标量,表示输入样本的分数,分数越高,输入图像就越有可能是原始的未压缩图像,反之,输入图像越可能是生成图像。

3 损失函数

3.1 生成器损失函数

G 的目标是删除 JPEG 压缩产生的伪影,恢复 JPEG 压缩图像并输出重建图像。为了使输出的重建图像在不牺牲图像视觉质量的情况下欺骗现有的 JPEG 压缩取证检测器,设计了损失函数 L_G,L_G 由逐像素损失 L_{pixel}、AC 分量损失 L_{acc}、校准损失 L_{cal} 和对抗损失 L_{adv} 构成,具体如式(1)所示,其中 α、β、η 和 γ 为每个损失的权重。

$$L_G = E_{X'}[\alpha L_{pixel} + \beta L_{acc} + \eta L_{cal} + \gamma L_{adv}] \tag{1}$$

L_{pixel} 反应原始图像和 G 输出图像之间的像素差异,其中 x_i 为原始未压缩图像,$G(x')$ 为 G 生成的重建图像,$S=W\times H$ 为图像的大小,如式(2)所示:

$$L_{pixel} = \frac{1}{S}\sum_{i=1}^{S} \| x_i - G(x') \|_2^2 \tag{2}$$

L_{acc} 恢复由 JPEG 压缩造成的高频分量损害,提高重建图像的视觉质量,如式(3)所示,其中 $D_{DCT}(\cdot)$ 表示 8×8 块的 DCT 变换,$\boldsymbol{M}$ 是一个掩码矩阵,如式(4)所示,$\overline{D_{DCT}(x^k)\odot\boldsymbol{M}}$ 为 $D_{DCT}(x^k)\odot\boldsymbol{M}$ 的平均值,$\odot$ 表示矩阵的点积。

$$L_{acc} = \frac{1}{N}\sum_{k=1}^{N} \| (D_{DCT}(x^k)\odot\boldsymbol{M} - \overline{D_{DCT}(x^k)\odot\boldsymbol{M}}) - (D_{DCT}(G(x')^k)\odot\boldsymbol{M} - \overline{D_{DCT}(G(x')^k)\odot\boldsymbol{M}}) \|_2^2 \tag{3}$$

$$\boldsymbol{M} = \begin{bmatrix} 0 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & & \vdots & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix} \tag{4}$$

L_{cal} 用于最小化生成图像与校准版本之间高频子带的方差差距,实现反取证的性能,其中 $D_{block_dct_j}(\cdot)$ 表示图像 8×8 DCT 变换后提取的第 j 个高频子带的所有 DCT 系数,$V_{val}(\cdot)$ 是输入向量的方差,如式(5)所示:

$$L_{cal} = \frac{1}{28}\sum_{j=1}^{28} | V_{val}(D_{block_dct_j}(G(x'))) -$$

$$V_{\mathrm{val}}(D_{\mathrm{block_det}_j}(G(x')_{\mathrm{cal}}))\mid \tag{5}$$

为了使得 G 生成图像与原始图像的统计特征相似,可以欺骗 D,获得更好的反取证性能,添加对抗损失 L_{adv},对应的损失函数具体如式(6)所示:

$$L_{\mathrm{adv}}=-D(G(x')) \tag{6}$$

3.2　鉴别器损失函数

D 负责判别输入图像是生成的重建图像还是原始未压缩图像。在 D 中设计一个损失函数,如式(7)所示,其中 $D(\cdot)$ 和 $G(\cdot)$ 分别代表鉴别器的输出和生成器的输出。L_{D} 值越小,说明生成的重建图像与原始未压缩图像的分布越近。

$$L_{\mathrm{D}}=-E_x[M_{\min}(0,-1+D(x))]-E_{x'}[M_{\min}(0,-1-D(x'))] \tag{7}$$

4　仿真实验与结果分析

4.1　实验配置

实验运行的环境为 Windows10 的 64 位操作系统,并采用 NVIDIA RTX 3080Ti GPU 显卡加速图像处理,编程语言为 Python,深度学习平台为 Tensorflow。在实验过程中,使用 3 个公共图像数据集:BossBase V1.01(BossBase)[13]、BOWS2-Original(BOWS)[14] 和 UCID-V2(UCID)[15]。从 BossBase 和 BOWS 数据集中随机抽取 16 000 张未压缩图像,每个图像都被随机裁剪成 9 个不重叠的 128×128 张图像作为训练数据集,共有 144 000 张 128×128 训练集图像。将 BossBase 和 BOWS 数据集中剩下的 4 000 张大小为 512×512 图像作为测试图像。

在训练过程中,模型训练迭代超过 85 000 次,前 10 000 次迭代只训练 G,后 75 000 次迭代 D 和 G 交替训练,设置批量大小为 16,固定学习率为 1.0×10^{-4},将式(1)中损失函数的权重分别设置为 $\alpha=1.0$,$\beta=1.0$,$\eta=50$,$\gamma=1.0\times10^{-5}$,默认采用 Adam 优化器。

4.2　图像质量的评估指标

该方法使用两种评价指标对实验结果进行定量对比,分别为结构相似性(SSIM)和峰值信噪比(PSNR)。SSIM 用于衡量两幅图像之间的相似度,通过亮度、对比度和结构 3 个方面来评估两幅图像的相似性,SSIM 值的范围在[0,1],数值越大表示图片质量越好。而 PSNR 是一种衡量图像质量的度量标准,通过计算有效信息与噪声之间的比率来反映图像是否失真,PSNR 值越大表示图像的质量越好。对应的计算公式分别如式(8)式(9)所示,其中分别用 P_{PSNR} 和 S_{SSIM} 表示 PSNR 和 SSIM 值。

$$S_{\mathrm{SSIM}}(x,y)=[I(x,y)^{\alpha}\cdot c(x,y)^{\beta}\cdot s(x,y)^{\gamma}] \tag{8}$$

$$P_{\mathrm{PSNR}}=10\cdot\ln\left(\frac{P_{\mathrm{MAX}}{}^2}{E_{\mathrm{MS}}}\right) \tag{9}$$

在式(8)、式(9)中,$I(x,y)$、$c(x,y)$ 和 $s(x,y)$ 分别为亮度对比函数、对比度对比函数和结构对比函数,E_{MS} 为均分误差,P_{MAX} 表示图像可能的最大像素值。

4.3　实验结果分析

为了评估 RBAM-JAF 模型生成图像的视觉质量,采用峰值信噪比和结构相似性这两个指标来评估生成图像的视觉质量并以原始未压缩图像作为参考,将测试集中的 4 000 张图像进行 JPEG 压缩,压缩图像质量分别为 $Q_{\mathrm{F}}=25$、50 和 75,并使用它们对应的未压缩图像作为参考,计算它们的平均 PSNR 和 SSIM 值,分别用 $P_{\mathrm{avg_PSNR}}$ 和 $S_{\mathrm{avg_SSIM}}$ 表示。此外,将该方法分别与两种传统的 JPEG 压缩反取证方法和两种基于 GAN 的 JPEG 压缩反取证方法进行比较:Stamm[4]、Fan[5]、Luo[1] 和 Wu 等[6],记为 M_1、M_2、M_3 和 M_4,RBAM-JAF 方法记为 M_5,J 记为 JPEG 压缩。为了确保公平比较,在实验过程中对所有方法采用了相同的设置。

从图 4 可以看出:JPEG 压缩后的图像具有明显的块状伪影,而反取证生成的图像具有更清晰的纹理和更清晰的边缘,压缩伪影和振铃效应大大降低,使得生成的图像在视觉上更接近原始未压缩的图像。表 1 所示:方法 M_5 与 M_1、M_2、M_3 和 M_4 相比,在大多数情况下具有最高的 PSNR 和 SSIM 值,因此,所生成的图像质量与这 4 种最先进反取证方法相比更好。此外,方法 M_5 在 JPEG 压缩的任何 Q_{F} 值下均获得了 PSNR 和 SSIM 的最佳性能。

(a) 原始图像　(b) JPEG 压缩图像　(c) 反取证生成的图像

图 4　JPEG 反取证方法实例图像

Fig. 4　Example image by JPEG anti-forensics method

为了评估方法 M_5 的反取证性能,将方法 M_5 与现有的 4 种反取证方法 M_1、M_2、M_3 和 M_4 的反取证性能作对比。实验过程中,将 4 000 张反取证重建图像作为正样本,对应的 4 000 张原始未压缩图像作为负样本,用两种

基于传统的 JPEG 取证检测器和两种基于 CNN 的检测器进行检测,分别为 Fan[16]、Val[17]、Yang[18] 和 Singh[19] 所提出的取证检测器。对每种反取证方法进行检测,分别用 F_1、F_2、F_3 和 F_4 表示,并计算 3 种 Q_F 值的 ROC 曲线下面积(AUC),若 AUC 值接近或小于 0.5,则反取证方法可以有效地欺骗取证检测器。如表 1 所示,反取证方法 M_5 的 AUC 值在基于传统和 CNN 的取证检测器下均接近或小于 0.5,能够有效地欺骗 4 种,取证检测器,与反取证方法 M_2 具有相当的反取证性能,且与方法 M_1、M_3 和 M_4 相比,方法 M_5 的反取证性能优于这 3 种反取证方法。

综上所述,方法 M_5 相比 M_1、M_2、M_3 和 M_4 4 种反取证方法,在生成图像质量和反取证的性能方面均取得了更优的结果。因此,方法 M_5 能够在不牺牲图像视觉质量的情况下欺骗现有的取证检测器。

表 1 各种算法的图像视觉质量和反取证性能结果对比

Table 1 Comparison of image visual quality and anti-forensics performance results of various algorithms

Q_F	方 法	AUC 值					
		F_1	F_2	F_3	F_4	P_{avg_PSNR}	S_{avg_SSIM}
25	J	0.987	0.971	1.000	1.000	32.580	0.886
	M_1	0.810	0.885	0.927	0.156	30.762	0.807
	M_2	**0.501**	**0.497**	**0.338**	**0.269**	32.312	0.889
	M_3	0.910	0.925	0.983	**0.684**	32.883	0.894
	M_4	0.752	0.486	0.939	0.717	33.359	0.906
	M_5	**0.668**	**0.504**	**0.453**	**0.619**	**33.382**	**0.911**
50	J	0.965	0.982	1.000	1.000	35.067	0.929
	M_1	**0.651**	0.904	0.722	0.173	32.006	0.855
	M_2	**0.498**	**0.511**	**0.316**	**0.287**	34.515	0.924
	M_3	0.821	0.866	0.993	**0.693**	34.970	0.932
	M_4	0.723	**0.534**	0.972	0.712	35.209	0.936
	M_5	**0.496**	**0.274**	**0.355**	**0.637**	**35.277**	**0.938**
75	J	0.921	0.995	1.000	1.000	37.756	0.954
	M_1	**0.503**	0.896	**0.487**	0.196	32.119	0.868
	M_2	**0.518**	**0.664**	**0.317**	**0.334**	36.679	0.947
	M_3	**0.674**	0.902	0.947	0.704	36.593	0.951
	M_4	**0.623**	0.928	0.929	0.726	36.890	0.957
	M_5	**0.451**	**0.603**	**0.285**	**0.646**	**37.983**	**0.962**

5 结 论

针对 JPEG 反取证方法在生成重建图像质量与反取证性能之间平衡不足的问题,设计了一种多级残差块和通道注意力机制相结合的反取证方法,在多级残差块的末端引入通道注意力机制,使 RBAM-JAF 专注于更有用的通道信息,所生成的图像能够更好地保持图像的视觉质量,并有效地欺骗现有的取证检测器。此外,在该模型中引入了特征融合模块,使其充分利用每个卷积层的特征信息,提高了生成图像的质量。实验结果表明:RBAM-JAF 模型不仅提高了生成图像的质量,还能够有效欺骗现有取证检测器,与现有的反取证方法相比,该方法在图像视觉质量和反取证性能方面均具有显著优势,并在恢复图像质量和反取证性能之间取得了良好的平衡。未来,将致力于研究如何在保持生成图像质量的情况下,提高反取证性能。

参考文献(References):

[1] LUO Y, ZI H, ZHANG Q, et al. Anti-forensics of JPEG compression using generative adversarial networks[C]//2018 26th European Signal Processing Conference (EUSIPCO). IEEE, 2018: 952-956.

[2] 高玉晓. 浅谈 JPEG 压缩反取证技术未来研究思路[J]. 信息系统工程, 2020(6): 12-13.
GAO Yu-xiao. Discussion on the future research ideas of JPEG compression anti-forensics technology[J]. China CIO News, 2020(6): 12-13.

[3] STAMM M C, TJOA S K, LIN W S, et al. Anti-forensics of JPEG compression[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2010: 1694-1697.

[4] STAMM M C, LIU K J R. Anti-forensics of digital image compression[J]. IEEE Transactions on Information Forensics and Security, 2011, 6(3): 1050-1065.

[5] FAN W, WANG K, CAYRE F, et al. JPEG anti-forensics with improved tradeoff between forensic undetectability and image quality[J]. IEEE Transactions on Information Forensics and Security, 2014, 9(8): 1211-1226.

[6] WU J, LIU L, KANG X, et al. A generative adversarial network framework for JPEG anti-forensics[C]//2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2020: 1442-1447.

[7] WU J, KANG X, YANG J, et al. A framework of generative adversarial networks with novel loss for JPEG restoration and anti-forensics[J]. Multimedia Systems, 2021, 27(6): 1075-1089.

[8] 何沛松, 李伟创, 张婧媛, 等. 面向 GAN 生成图像的被动取证及反取证技术综述[J]. 中国图象图形学报, 2022, 27(1): 88-110.
HE Pei-song, LI Wei-chuang, ZHANG Jing-yuan, et al. Overview of passive forensics and anti-forensics techniques for GAN-generated image[J]. Journal of Image and Graphics, 2022, 27(1): 88-110.

[9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 770-778.

[10] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.

[11] 储岳中, 乔雨楠. 多注意力结合光流的视频超分辨方法[J]. 重庆工商大学学报(自然科学版), 2022, 39(4): 1-8.
CHU Yue-zhong, QIAO Yu-nan. Video super-resolution method of multi-attention combined with optical flow[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022, 39(4): 1-8.

[12] SETIADI D R I M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography[J]. Multimedia Tools and Applications, 2021, 80(6): 8423-8444.

[13] BAS P, FILLER T, PEVNÝ T. Break our steganographic system: The ins and outs of organizing BOSS[C]//Information Hiding. Berlin, Heidelberg: Springer, 2011: 59-70.

[14] 赵鑫, 王垚飞, 陈可江, 等. 引入超分辨率下采样误差的图像边信息估计隐写[J]. 中国图象图形学报, 2022, 27(1): 226-237.
ZHAO Xin, WANG Yao-fei, CHEN Ke-jiang, et al. Spatial image steganography based on side information estimated by super resolution[J]. Journal of Image and Graphics, 2022, 27(1): 226-237.

[15] SCHAEFER G, STICH M. UCID: An uncompressed color image database[C]//Proceedings of the Storage and Retrieval Methods and Applications for Multimedia 2004, 2003: 472-480.

[16] FAN W, WANG K, CAYRE F, et al. A variational approach to JPEG anti-forensics[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 3058-3062.

[17] VALENZISE G, TAGLIASACCHI M, TUBARO S. Revealing the traces of JPEG compression anti-forensics[J]. IEEE Transactions on Information Forensics and Security, 2013, 8(2): 335-349.

[18] YANG C, LI H, LIN F, et al. Constrained R-cnn: a general image manipulation detection model[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. IEEE, 2020: 1-6.

[19] SINGH G, GOYAL P. SDCN2: A Shallow Densely connected CNN for multi-purpose image manipulation detection[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2022, 18(2): 1-22.

责任编辑:李翠薇