计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(7): 1786−1796, 2025

基于窗口自注意力特征融合的感知图像认证哈希

周元鼎　高国鹏　房耀东　秦　川
（上海理工大学光电信息与计算机工程学院　上海　200093）
（221240081@st.usst.edu.cn）

Perceptual Image Authentication Hashing Based on Window Self-Attention Feature Fusion

Zhou Yuanding, Gao Guopeng, Fang Yaodong, and Qin Chuan
（*School of Optical-Electrical and Computer Engineering*, *University of Shanghai for Science and Technology*, *Shanghai* 200093）

Abstract　With the rapid development of multimedia and network technology, the security of digital image content is becoming more and more prominent. In this paper, we propose a deep perceptual image authentication hashing scheme based on window self-attention feature fusion, which can effectively detect whether the perceptual content of the original image has changed. It can be applied to content authentication, tampering recognition, copy detection, and other similar scenarios. This scheme uses a convolutional neural network architecture that integrates a window self-attention mechanism to build a hashing model that encompasses global and local image features. The scheme chunks the shallow features obtained from the backbone network and extracts the corresponding window features, then calculates the correlation between each intermediate local feature and the global feature to filter out the final local features, and finally inputs the local features and global features into the Hash generation module for fusion and compression to obtain the final image Hash code. In the training process, an integrated loss function based on Hash loss and classification loss is used to constrain the model to improve the robustness and discrimination. The experimental results show that this scheme can achieve superior image content authentication performance compared with existing typical perceptual authentication hashing schemes.

Key words　perceptual hashing; image authentication; window self-attention; feature fusion; robustness

摘　要　随着多媒体和互联网技术的快速发展，数字图像内容的安全性问题日益突出. 为此，提出了一种基于窗口自注意力特征融合的深度感知图像认证哈希方案，该方案能有效检测原始图像的感知内容是否发生变化，并可应用于内容认证、复制检测、篡改识别等场合. 该方案以卷积神经网络为基础，利用窗口自注意力构建了一个融合图像全局和局部特征的哈希模型. 模型首先对主干网络获得的浅层特征进行分块并提取相应的窗口特征，然后计算每个局部特征与全局特征之间的相关性来筛选出最终的局部特征，再将这部分特征和全局特征输入到哈希生成模块中进行融合与压缩，得到最终的图像哈希码. 在训练过程中，利用哈希损失和分类损失构造的联合损失函数对模型进行约束，提高感知认证哈希方案的鲁棒性和唯一性. 实验结果表明，与现有典型的感知认证哈希方案相比，所提方案可获得更优的图像内容认证性能.

关键词　感知哈希；图像认证；窗口自注意力；特征融合；鲁棒性

中图法分类号　TP391

收稿日期：2023-08-21；**修回日期**：2025-01-09

基金项目：国家自然科学基金项目（62172280, U20B2051）
This work was supported by the National Natural Science Foundation of China (62172280, U20B2051).

通信作者：秦川（qin@usst.edu.cn）

DOI: 10.7544/issn1000-1239.202330669 CSTR: 32373.14.issn1000-1239.202330669

互联网为内容创作者提供了一个低成本、高效率的平台，使他们能够将创作内容快速传播到各个社交平台. 同时，随着 Twitter、Instagram 和微博等社交媒体的兴起，用户可以通过这些平台方便快捷地与他人分享数字图像. 然而，随着多媒体技术的快速发展，社交平台上的数字图像数据量不断增加，这大大增加了图像被篡改和伪造的概率. 此外，一些操作简单的图像编辑工具的出现，使得没有专业技能的人也可以轻松地对图像进行修改和编辑. 因此，如何验证数字图像的真实性已经成为一个重要的问题[1]. 近期，研究人员提出了许多基于水印[2-3]和取证[4-5]的图像内容认证方法. 图像水印需要将水印信息事先嵌入到图像中，通过分析提取后的水印信息来实现内容认证，但嵌入操作在一定程度上破坏了图像的质量和完整性. 图像取证虽然无需嵌入信息，但它涉及复杂的计算过程并且认证的准确性也相对较低. 感知认证哈希（perceptual authentication hashing, PAH）作为一种半主动的认证方法，可通过比对图像间的哈希距离来实现认证，近年来引起了研究人员的广泛关注.

需要指出的是，MD5 和 SHA-1 等密码学哈希函数虽然也可以通过将任意长度的数据映射成固定长度的哈希码来实现数据的完整性认证和篡改检测，但由于这些密码学哈希函数对输入数据的变化特别敏感，故不适用于数字图像这种多媒体数据的内容认证. 数字图像作为在互联网上广泛传播的多媒体信息，难免会经历一些图像处理操作（如 JPEG 压缩、图像增强等），但这些操作并不会改变图像的感知内容. 故对这些处理后的图像，虽然数据表示上有所改变，但要求哈希函数仍应生成和原始图像相同或者相近的哈希值，且不能判断图像被篡改. 所以，对输入敏感的密码学哈希函数显然无法用于图像内容认证的实际场合. 而 PAH 利用图像的感知内容特征来生成紧凑的哈希码，可以很好地解决传统密码学哈希的这个问题，已被广泛应用于篡改识别[6-7]、图像认证[8]和拷贝检测[9-10]等领域.

PAH 侧重于图像内容的认证，该任务中的感知相似图像一般是指原始图像经过内容保留操作得到的图像，其感知内容和原图是相同的. 一个理想的 PAH 方案需要同时满足 2 个基本条件[11-12]：感知鲁棒性和唯一性. 感知鲁棒性要求 PAH 方案对感知相似的 2 幅图像生成相同或相近的图像哈希码，即该方案能够有效地抵抗图像内容保留操作. 唯一性，也称抗冲突性，要求 PAH 方案可以区分感知不同的图像，即对于感知不相似的图像生成不同的哈希码. 有些 PAH 方案还需要考虑密钥依赖性，即在不同密钥的条件下生成显著不同的哈希码.

在 PAH 研究的早期，许多基于传统图像处理的方案被提出. 这些方案一般包含 3 个主要步骤：图像预处理、特征提取和哈希生成. 在特征提取步骤中，通常采用人工设计的方式来提取图像特征. 随着深度学习的不断发展，卷积神经网络（convolutional neural network, CNN）作为一种强大的特征提取器被广泛应用于图像处理和计算机视觉领域. 例如，研究人员成功地将 CNN 应用于语义检索哈希[13]，实现了大规模图像检索. 然而，在图像内容认证方面，基于深度学习的 PAH 研究相对较少，且现有的方案涉及的图像内容保留操作较少，提取的特征也相对单一. 因此，为进一步提高图像内容认证的性能，本文提出一种新的基于 CNN 的深度 PAH 方案，其主要贡献有 3 方面：

1）提出一种基于窗口自注意力的特征融合方案，通过计算图像窗口特征与全局特征之间的相关性，筛选与全局特征相关性最大的若干窗口特征作为局部特征，融合局部特征与全局特征生成哈希码.

2）增加了产生感知相似图像的内容保留操作种类，构建了一个针对 PAH 的数据集用来训练和测试；同时设计了联合损失函数用来提高模型的鲁棒性和唯一性.

3）本文算法与现有的 PAH 方案相比，在图像的内容认证方面有更好的性能表现，可有效应用于图像近似复制检测和篡改认证.

1 相关工作

1.1 感知图像认证哈希

现有的感知图像认证哈希方案可以分为 2 类：基于传统图像处理的方案和基于深度学习的方案. 基于传统图像处理的感知图像认证哈希方案需要通过手动设计的方式来提取图像特征并生成哈希码. 基于深度学习的方案则是通过网络模型和大规模数据集的训练，学习得到图像的感知特征并生成哈希码.

基于传统图像处理的感知图像认证哈希方案，依据特征提取方式可以细分为 3 类：1）基于特征点的方案. 这类方案通过使用一些具有代表性的图像显著特征点提取器来提取鲁棒特征. Ouyang 等人[14]

使用四元数 Zernike 矩和 SIFT 算法分别提取图像的全局特征和局部特征以生成图像哈希，其中全局特征可以用来进行图像的内容认证，而局部特征可以用来进行图像篡改区域的定位以及篡改操作的识别. Singh 等人[15]利用 KAZE 特征点提取方法来获取图像的局部特征，并利用奇异值分解获取图像的全局特征，将这 2 个特征进行简单的拼接并通过随机化的方法得到最终的哈希码. 2)基于统计特征的方案. 这类方案利用图像分析中重要的统计特征来进行感知特征的提取. Srivastava 等人[16]为了有效捕捉图像的视觉感知内容，使用均值、方差、偏度和峰度这 4 个统计特征来设计感知哈希方案，该方案对常见的数字图像处理操作 (除旋转外) 有较好的鲁棒性. 3)基于变换域的方案. 这类方案在变换域中提取图像有效的特征信息. Qin 等人[17]利用离散傅里叶变换对预处理后的图像进行鲁棒频率特征的提取，利用非均匀采样的方法从中低频分量中选择更多的采样点，并对采样特征进行增加扰动和量化处理来安全地产生二进制哈希码. Abdullahi 等人[18]将傅里叶-梅林变换和分形编码相结合用于生成高效的图像哈希码.

基于深度学习的感知图像认证哈希方案主要是利用深度神经网络模型，从训练集中学习图像的鲁棒表示. Li 等人[19]提出一种基于数据驱动的感知图像哈希方案，通过训练降噪自编码器寻找到图像与哈希码之间的最佳映射关系. Qin 等人[20]提出一种多约束的深度感知图像哈希模型，使用 CNN 作为特征提取器，并通过构建相似和不相似的图像及特征图谱间的对称约束来提高模型的认证性能.

1.2 注意力机制

近年来，注意力机制已成为捕捉数据特征依赖关系的重要模型组件之一，尤其是自注意力机制，它可以对特征序列中所有位置的向量进行加权平均，来计算每个位置对整体特征的依赖程度. 注意力机制最初多应用于自然语言处理领域，例如机器翻译、文本生成等. 后来在计算机视觉领域中，研究人员开始探索利用注意力机制来获取图像特征间的依赖关系，以提升计算机视觉分析和处理的准确性. 例如，Zhang 等人[21]利用自注意力机制获取图像间的依赖关系，从而实现图像合成；Wang 等人[22]使用注意力机制来捕获图像局部特征以进行细粒度的图像分类；Cao 等人[23]利用注意力模块提取图像的局部特征，并用于将全局特征检索出的图像与查询图像之间的局部特征点进行匹配，以提升检索的准确率.

2 本文方法

本文旨在构建一个端到端的深度感知图像认证哈希模型用于图像内容认证. 该模型利用 CNN 强大的特征学习能力，提取图像的全局特征和局部特征并进行融合，丰富用于生成哈希码的图像表征. 本文提出的 PAH 网络模型主要包括卷积主干网络、全局特征提取模块、局部特征提取模块和哈希生成模块 4 个部分. 图 1 展示了该深度 PAH 模型的具体架构. 模型的输入为原图、相似图像、不相似图像所构成的三元组. 局部特征提取模块利用窗口自注意力机制对图像进行分块，并提取图像块的特征表示. 哈希

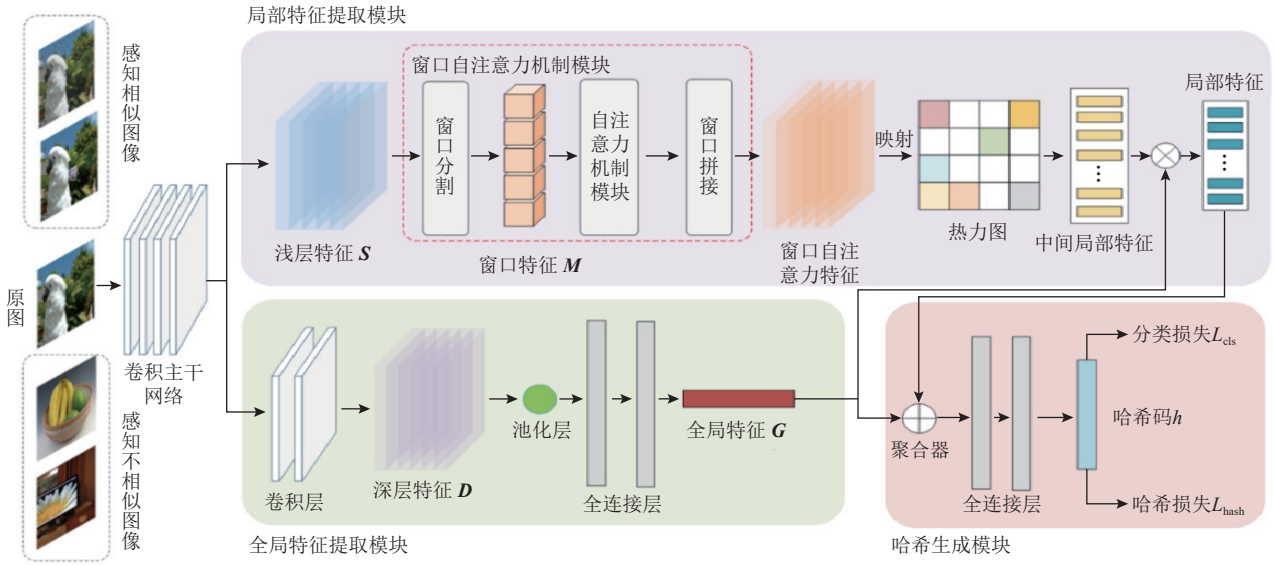


Fig. 1 Network architecture of our PAH method

图 1 本文 PAH 方法的网络架构

生成模块进行特征融合，并通过全连接层压缩特征生成最终的哈希码. 另外，哈希生成模块的输出还用于构建联合损失，指导深度 PAH 模型训练.

2.1 卷积主干网络和哈希生成模块

卷积主干网络模块的作用是用来对输入图像进行预处理和初步特征提取，为全局特征和局部特征提取模块提供便利. 该网络包含了 8 个卷积层和 1 个最大池化层，其中第 1 个卷积层采用 7×7 的卷积核，结合最大池化层对输入图像进行降维，同时增加特征的通道数，以保留有效的特征信息. 例如，对于一个尺寸为 224×224×3 的图像，经过主干网络的第一个卷积层和池化层处理后，输出维度为 56×56×64 的图像.

哈希生成模块如图 1 右下角所示，由聚合器和全连接层组成. 聚合器的作用是将全局特征和局部特征提取模块提取的特征进行拼接；全连接层的作用是对拼接后的特征进行融合和压缩. 哈希生成模块中最后一个全连接层输出深度 PAH 模型的最终哈希码，其输出长度为 l.

2.2 全局特征与局部特征提取模块

如图 1 所示，对于给定的一张图像，经过卷积层后可得到 2 个特征图 $\boldsymbol{S}\in\mathbb{R}^{H_S\times W_S\times C_S}$ 和 $\boldsymbol{D}\in\mathbb{R}^{H_D\times W_D\times C_D}$，分别代表浅层特征和深层特征，其中 H, W, C 分别表示特征图的高度、宽度和通道数. 一般来说，深层特征 $\boldsymbol{D}$ 在空间上的特征尺寸较小，但通道数较多[23]，即 $H_D\leqslant H_S$, $W_D\leqslant W_S$, $C_D\geqslant C_S$. 浅层特征 $\boldsymbol{S}$ 用于提取局部特征，而深层特征 $\boldsymbol{D}$ 用于生成全局特征.

全局特征提取模块对卷积层提取到的深层特征 $\boldsymbol{D}$ 进行进一步处理，利用全局平均池化对 $\boldsymbol{D}$ 中的每个特征图取平均值，将特征变成一个维度为 1×1×C 的特征向量. 全局平均池化可以大幅度降低网络的参数值，同时防止模型过拟合. 另外，再通过全连接层 F 对得到的特征进行整合以生成全局特征 $\boldsymbol{G}$:

$$\boldsymbol{G}=F\times\left[\frac{1}{H_DW_D}\sum_{h,w}\boldsymbol{d}_{h,w}^{(1)},\cdots,\frac{1}{H_DW_D}\sum_{h,w}\boldsymbol{d}_{h,w}^{(k)}\right],\tag{1}$$

其中 $\boldsymbol{d}_{h,w}^{(k)}\in\mathbb{R}^{C_D}$ 表示深层特征 $\boldsymbol{D}$ 中第 k 个特征图在 (h, w) 位置上的特征.

局部特征提取模块包含窗口自注意力模块和局部特征生成模块 2 部分. 图 2 以 3 通道彩色图像作为窗口自注意力模块的输入，展示了窗口分割和自注意力机制的处理流程. 窗口分割对输入的浅层特征 $\boldsymbol{S}$ 按照预先设置的窗口大小 B 进行分割，得到相应的窗口特征 $\boldsymbol{M}=(\boldsymbol{M}_1,\boldsymbol{M}_2,\cdots,\boldsymbol{M}_Q)\in\mathbb{R}^{Q\times H_M\times W_M\times C_M}$，其中 Q 表示窗口的数量，$\boldsymbol{M}_i$ 表示第 i 个窗口特征. 自注意力机制对每个窗口特征 $\boldsymbol{M}_i$ 使用 1×1 的卷积进行转换，得到 3 个相应的特征 $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{V}_3$. $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$ 被用于联合生成自注意力特征图 $\boldsymbol{E}\in\mathbb{R}^{H_M\times W_M}$:

$$\boldsymbol{E}=\sigma(\boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{V}_2),\tag{2}$$

其中 $\sigma(\cdot)$ 表示 Softmax 激活函数. $\boldsymbol{V}_3$ 与 $\boldsymbol{E}$ 经过聚合和卷积变换后得到窗口自注意力图. 窗口拼接部分将生成的多个窗口自注意力特征图按照窗口分割的顺序进行特征的拼接，并利用映射关系得到包含 Q 个窗口信息的中间局部特征 $\boldsymbol{\mathcal{T}}\in\mathbb{R}^{N\times Q\times C}$.

为了选取出最重要的 k 个窗口特征，$k=\theta\times Q$(θ 表示中间局部特征筛选比例)，利用式(3)计算出全局特征 $\boldsymbol{G}\in\mathbb{R}^{N\times C}$ 与每个中间局部特征 $\boldsymbol{\mathcal{T}}_i\in\mathbb{R}^{N\times C}$ 之间的相关性[24]，并根据相关性程度排序筛选出相关性最大的 k 个中间局部特征来生成最终的局部特征.

$$c(\boldsymbol{\mathcal{T}}_i,\boldsymbol{G})=\langle\boldsymbol{\mathcal{T}}_i,\boldsymbol{G}\rangle,i=1,2,\cdots,Q.\tag{3}$$

2.3 损失函数

为了提高哈希生成的质量，需要对 PAH 模型学习到的特征进行进一步的优化. 损失函数在这环节中起到重要的作用，它可以约束模型朝着预先设定的方向进行学习，有效地保证了模型的内容认证性

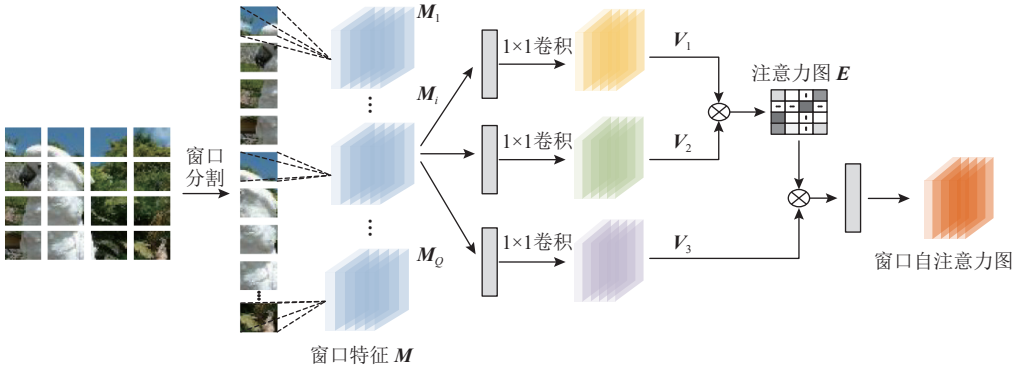


Fig. 2 Illustration of the window self-attention module

图 2 窗口自注意力模块示意图

能. 本文方法的损失函数主要由 2 部分组成, 即哈希损失 L_{hash} 和分类损失 L_{cls}, 两者联合来约束模型的鲁棒性和唯一性, 总损失函数如式(4)所示.

$$L=\lambda_1 L_{\text{cls}}+\lambda_2 L_{\text{hash}}, \tag{4}$$

其中 λ_1 和 λ_2 分别为分类损失和哈希损失的超参数.

本文深度 PAH 模型的输入是原图、相似图像、不相似图像的三元组图像形式, 可以用集合 $O=\{\boldsymbol{I}, \widehat{\boldsymbol{I}}_1, \widehat{\boldsymbol{I}}_2, \cdots, \widehat{\boldsymbol{I}}_N, \boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_N\}$表示, 其中 $\boldsymbol{I}$ 表示原始图像, $\widehat{\boldsymbol{I}}$ 表示与原始图像感知相似的图像, $\boldsymbol{P}$ 表示与原始图像感知不相似的图像, N 表示感知相似图像和不相似图像的数量, 图像输入的总数量为 $2N+1$.哈希损失 L_{hash} 的目的是减小原图和其感知相似图像之间的哈希距离, 同时扩大原图和其感知不相似图像之间的哈希距离. 哈希码间的距离可以通过均方误差进行衡量, 这样原图和其相似及不相似图像之间的距离函数可以表示为

$$L_{\text{sim}}^{(i)}=\frac{1}{l}\sum_{k=1}^{l}\left\|\mathcal{H}(\boldsymbol{I})^{(k)}-\mathcal{H}(\widehat{\boldsymbol{I}}_i)^{(k)}\right\|_2^2, \tag{5}$$

$$L_{\text{dif}}^{(i)}=\frac{1}{l}\sum_{k=1}^{l}\left\|\mathcal{H}(\boldsymbol{I})^{(k)}-\mathcal{H}(\boldsymbol{P}_i)^{(k)}\right\|_2^2, \tag{6}$$

其中 $\mathcal{H}(\cdot)$表示本文的深度 PAH 算法, 可以将图像映射到一个固定长度的哈希码; $\widehat{\boldsymbol{I}}_i$和 $\boldsymbol{P}_i$分别表示第 i 张感知相似和不相似图像; l 表示生成的哈希码的长度; $\mathcal{H}(\boldsymbol{I})^{(k)}$表示哈希码中第 k 个值. 然而, 直接使用均方误差得到的哈希码之间的距离相对较大, 如果将其作为损失直接传播到模型中, 将会导致模型难以收敛, 无法很好地优化. 因此, 需要对得到的距离做进一步的归一化处理, 以消除大尺度距离对模型的影响. 哈希损失 L_{hash} 由相似损失和不相似损失 2 部分构成, 即

$$L_{\text{hash}}=\frac{1}{N}\sum_{i=1}^{N}Sigmoid(L_{\text{sim}}^{(i)})-Sigmoid(L_{\text{dif}}^{(i)}), \tag{7}$$

其中 $Sigmoid(\cdot)$ 表示归一化函数, 它可以将数值归一化到 [0.5, 1], 即越小的值经过函数 $Sigmoid(\cdot)$ 之后将越接近 0.5, 越大的值将越接近 1.

哈希损失 L_{hash} 主要用于对图像对进行约束, 以控制感知相似和不相似图像对之间的哈希距离, 但是并没有充分考虑到图像集合 O 中所有的不相似图像. 为了进一步提高本文提出的 PAH 模型的性能, 我们引入了分类损失来对模型进一步约束. 具体而言, 将原图 $\boldsymbol{I}$ 和其不相似图像 $\boldsymbol{P}$ 中的每一张图像都视为一个单独的类别, 即将集合 O 中的图像分为 $N+1$ 个类别, 其中原图和其感知相似图像属于同一个类别. 通过对这些图像进行伪标签的设定, 构建了分类损失 L_{cls}:

$$L_{\text{cls}}=\frac{1}{m}\sum_{i=1}^{m}\left(-\boldsymbol{h}_i^{(y_i)}+\ln\sum_{j=1}^{m}\exp(\boldsymbol{h}_i^{(j)})\right), \tag{8}$$

其中 $\boldsymbol{h}_i^{(j)}$表示第 i 张图像生成的哈希码中的第 j 个值; y_i 表示图像真实标签 y 中的第 i 个标签值, 真实标签由 $N+1$ 个不同的整数构成; $m=2N+1$ 表示每个批次中输入的图像数量.

3 实验结果与分析

本文进行了大量的实验来验证本文所提各个算法模块的有效性, 所有的实验都是在搭载 Windows 10 系统、NVIDIA GeForce RTX2080Ti GPU、Inter® Core™ i9-10900X CPU 和 32 GB 内存的台式机上运行的, 并且所有的深度感知图像认证哈希模型都是基于 Pytorch 框架进行开发的.

3.1 实验设置

1)相关指标. 图像哈希码之间的距离可以使用均方误差进行计算, 具体公式为:

$$D(\boldsymbol{h}_1,\boldsymbol{h}_2)=\frac{1}{l}\sum_{k=1}^{l}\|\boldsymbol{h}_1^{(k)}-\boldsymbol{h}_2^{(k)}\|_2, \tag{9}$$

其中 $\|\cdot\|_2$表示欧几里得范数. 对于每一对图像, 如果它们之间的哈希距离不大于给定的阈值 T, 则可以判定这 2 张图像为感知相似图像, 否则为感知不相似图像. 因此可以使用错误接受率(FAR)和错误拒绝率(FRR)来评估 PAH 方案的认证性能, 具体公式为:

$$\begin{aligned}&FAR(T)=Pr\{D(\boldsymbol{h}_1,\boldsymbol{h}_2)\leqslant T|\boldsymbol{I}_1\text{和}\boldsymbol{I}_2\text{感知不相似}\},\\&FRR(T)=Pr\{D(\boldsymbol{h}_1,\boldsymbol{h}_2)>T|\boldsymbol{I}_1\text{和}\boldsymbol{I}_2\text{感知相似}\},\end{aligned} \tag{10}$$

其中 T 为阈值. 显然, FAR 和 FRR 的值越小, 模型的认证性能就越好. 另外, $F1$ 分数常被用来量化 PAH 模型内容认证的准确性, 其值越大表示模型的内容认证性能越好, 其计算方式为:

$$F1=\max_{T}\left\{\frac{2[1-FAR(T)]\times[1-FRR(T)]}{[1-FAR(T)]+[1-FRR(T)]}\right\}. \tag{11}$$

2)训练设置. 本文的数据集分成训练集和测试集 2 部分, 其中的图像均是来自 COCO 数据集[25], 且训练集和测试集的图像是没有重复的. 具体地, 训练集中的图像分为原图、感知相似图像和感知不相似图像 3 部分, 且感知相似图像和感知不相似图像的数量是一致的, 其中感知相似图像是对原图使用表 1 列出的 62 个内容保留操作生成的. 这样训练集共包含了 187 500 张图像, 其中有 1 500 张原始图像、93 000

(1 500×62)张感知相似图像和 93 000 张感知不相似图像. 测试集与训练集的结构类似, 共包含了 62 500 张图像, 其中有 500 张原始图像、31 000(500×62)张感知相似图像和 31 000 张感知不相似图像. 图像在输入到模型时, 尺寸会被归一化至 128×128×3. 在模型训练中, 使用了初始学习率为 0.01 的 SGD 优化器对模型的参数进行更新, 并采用 StepLR 机制对优化器的学习率进行动态调整; 每经历 50 个训练轮次, 优化器的学习率会被乘以 0.1. 模型训练总共进行 300 个轮次, 模型的批量大小设置为 125(1+2×62). 分类损失和哈希损失的超参数 λ_1 和 λ_2 分别设置为 0.05 和 1, 哈希码的长度 l 为 128.

Table 1 Image Content-Preserving Manipulations and Its Parameters

表 1 图像内容保留操作及其参数

类型	操作	参数	数量
几何操作	旋转	旋转角度：5°，10°，15°，30°，45°	5
	翻转	方向：水平，垂直，水平+垂直，垂直+水平	4
	缩放	缩放比例：0.2，0.4，0.6，2	4
增强操作	高斯噪声	方差：0.01，0.02，0.03，0.04，0.05，0.06，0.07，0.08	8
	JPEG 压缩	压缩因子：10，20，30，40，50，70	6
	高斯模糊	模糊半径：1，2，3，4，5，6，7，8，9	9
	亮度调整	亮度因子：0.6，0.7，0.8，0.9，1.1，1.2，1.3，1.4	8
	锐化	锐化因子：[0.1, 5] 内的随机数	1
	像素化	像素化比例：[0.2, 0.5] 内的随机数	1
	滤波	分类：中值滤波，细节滤波，平滑滤波，模式滤波	4
	色彩空间调整	色彩空间：HSV 色彩空间、YCbCr 色彩空间	2
编辑操作	边框叠加	边框颜色：白色，黑色； 边框宽度：0.01，0.05，0.1	6
组合操作	组合操作 1	亮度因子：随机；对比度因子：随机	1
	组合操作 2	JPEG 压缩因子：70；高斯噪声方差：0.007；高斯模糊半径：0.7；旋转角度：5°；亮度因子：0.9	1
	组合操作 3	JPEG 压缩因子：50；高斯噪声方差：0.007；高斯模糊半径：0.7；旋转角度：5°；对比度因子：1.1	1
	组合操作 4	JPEG 压缩因子：70；高斯噪声方差：0.007；高斯模糊半径：0.7；旋转角度：5°；像素化比例：随机	1

3.2 窗口大小与局部特征分析

为了分析本文提出的算法在不同窗口尺寸 wz 与不同中间局部特征 $\boldsymbol{T}$ 数量条件下的性能, 本节进行了一系列实验. 图 3 展示了不同窗口尺寸和不同中间局部特征数量下本文深度 PAH 模型的接收者操作特征(receiver operating characteristic, ROC)曲线.

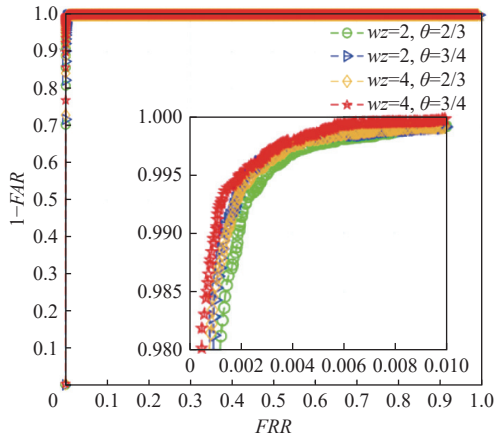


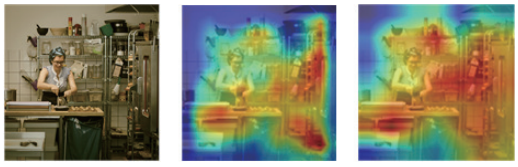
Fig. 3 ROC curves of the models with different parameter settings

图 3 不同参数设置下模型的 ROC 曲线

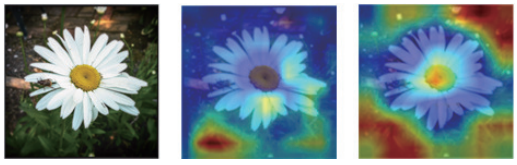
通过卷积主干网络得到图像的浅层特征 $\boldsymbol{S}$, 它的高度和宽度都是 4 的倍数. 这里我们设置了 2 种尺寸的窗口大小, 即 wz=2 和 wz=4, 用于窗口分割模块对 $\boldsymbol{S}$ 进行分割. 另外, 局部特征的数量是按照比例从窗口特征中进行挑选的, 本节中设置了 2 种比例 θ 进行分析, 分别是 θ=2/3 和 θ=3/4. 利用式(9)针对测试集中选取的相对于 500 张原图的 31 000 个感知相似图像对和 31 000 个感知不相似图像对, 即共 62 000 个图像对, 计算哈希距离, 再设置不同的阈值 T 通过式(10)计算相应的 FAR 和 FRR 值来绘制 ROC 曲线. ROC 曲线图的横坐标为 FRR、纵坐标为 1−FAR, 曲线越靠近左上角表明模型的性能越好. 由图 3 可以看出在这几种参数设置下模型都可以取得较好的认证性能, 相较而言, wz=4 和 θ=3/4 时, 模型的性能最好.

Grad-CAM[26] 是一种基于梯度的图像可视化技术, 它通过对目标图像的梯度进行全局平均池化来获取权重, 将权重与对应的特征图相乘并求和来生成热力图, 可用于深度学习模型关注于目标图像的

哪些区域的可视化. 为了更好地展现不同窗口尺寸 wz 下模型学习到的不同特征(θ=3/4), 本文利用 Grad-CAM 将模型学习到的特征以热力图的形式叠加到原始图像上, 如图 4 所示. 可以看出, 当窗口尺寸 wz=4 时, 本文的深度 PAH 模型可以关注到更多的区域. 需要注意的是, 相对于 SRH 模型主要关注图像中的目标对象信息, PAH 模型更加关注于图像的整体信息. 由图 4(d)~(f)看出, 对于一张类别为花的图像, 模型在关注“花”这个主体信息的同时, 也关注到了周围背景的内容. 但当窗口尺寸过大时, 窗口分割后可能只包含 1 个到 2 个窗口特征, 故本方案中窗口尺寸设置为 4.



(a) 原图1 (b) wz=2 (c) wz=4



(d) 原图2 (e) wz=2 (f) wz=4

Fig. 4 Characteristic heat maps at different window sizes

图 4 不同窗口尺寸下的特征热力图

3.3 鲁棒性分析

理想的 PAH 方案应该可以有效抵抗内容保留操作对图像哈希码的影响, 也就是说 PAH 方案要求原图和其感知相似图像生成的哈希码之间的距离要尽可能的小, 即鲁棒性强. 首先从 UCS-SIPI 图像数据集中选择了 6 张尺寸大小为 512×512 的标准测试图像, 如图 5 所示.

图 6 展示了测试图像在经过不同内容保留操作后的哈希距离, 其中横坐标表示不同内容保留操作



(a) 飞机 (b) 房屋 (c) 山魈



(d) 辣椒 (e) 帆船 (f) 树木

Fig. 5 Standard test images in the UCS-SIPI image dataset

图 5 UCS-SIPI 图像数据集中的标准测试图像

的参数名称, 纵坐标表示原图与其感知相似图像之间的哈希距离. 由于表 1 中像素化和锐化操作只有 1 个参数, 故它们的哈希距离并未通过曲线图展示. 从图 6 中可以观察到, 这些感知相似图像之间的哈希距离都小于 0.5, 且滤波操作、JPEG 压缩、高斯噪声、高斯模糊和缩放操作下得到的哈希距离都接近于 0, 即经过内容保留操作后, 这些图像在本文的深度 PAH 模型下生成的哈希码基本不变.

为了进一步展示本文模型的感知鲁棒性, 从实验测试集中选取了相对于 500 张原始图像的 31 000 个感知相似图像对. 表 2 列出了这些相似图像对之间哈希距离的统计值, 可以看出, 这 3 类内容保留操作的哈希距离均值都远小于 0.05, 组合操作下哈希距离的均值也达到了 0.050 7, 并且标准差也都小于 0.3. 值得注意的是, 增强操作类别中的大部分内容保留操作对应的哈希距离为 0, 表明使用这些操作得到的相似图像与原始图像的哈希码是完全一致的. 综合以上 2 个实验结果, 可以说明本文所提出的深度 PAH 模型具有良好的感知鲁棒性.

3.4 唯一性分析

唯一性是指 PAH 方案可以有效区分感知不相似的图像, 即对于 2 张感知不相似的图像, PAH 方案需生成完全不同的哈希码, 且使得它们之间的距离尽可能的大. 为了验证本文 PAH 方案的唯一性, 本节从 DISC21 数据集[27] 中挑选了 1 250 张不同场景、不同类别的图像来构建唯一性实验数据集. 该数据集中的图像两两不相似, 共包含 780 625(1 250×(1 250−1)/2)个感知不相似图像对. 因为感知哈希的唯一性和鲁棒性是 2 个相互制约的性能指标, 所以我们在设置不同阈值 T 的条件下来联合分析本文 PAH 方案的唯一性和鲁棒性.

表 3 列出了不同阈值 T 下, 本文 PAH 方案对于感知相似图像和感知不相似图像判断的正确率, 其中 $\Gamma_{\mathrm{sim}}(\Gamma_{\mathrm{dif}})$ 表示本文方案将实验数据集中的感知相似(感知不相似)图像对正确判断为感知相似(感知不相似)的比例, 即 Γ_{sim} 和 Γ_{dif} 分别对应方案的鲁棒性和唯一性. 从表 3 可以发现, 当阈值 T=0.9 时, Γ_{dif} 达到 97.10%, 同时从表 2 可以发现鲁棒性实验中绝大部分哈希距离都远小于 0.9, 当表 3 中 T=0.9 时, Γ_{sim} 达到了 97.14%. 这表明本文 PAH 方案可以有效地区分感知不相似图像和感知相似图像, 具有较好的唯一性且也是以 0.9 作为阈值.

3.5 性能比较

为了验证本文所提深度 PAH 方案的优越性,

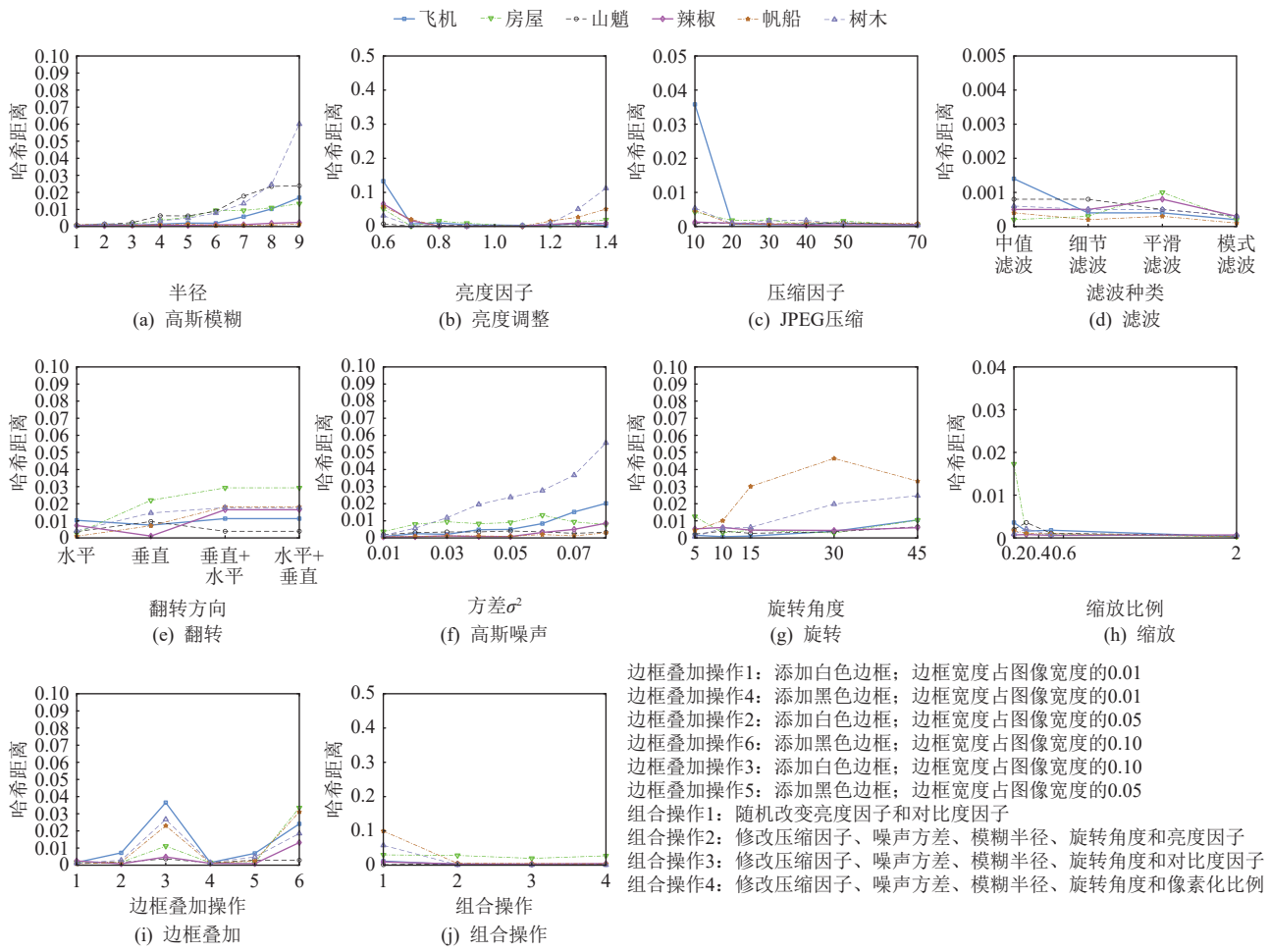


Fig. 6 Hash distance curve diagrams of six test images under different content-preserving manipulations

图 6 6 张测试图像在不同内容保留操作下的哈希距离曲线图

Table 2 Statistics of Hash Distance Between Perceptual Identical Images

表 2 感知相似图像之间哈希距离的统计

操作	最大值	最小值	均值	标准差
旋转	2.592 3	0.000 1	0.036 2	0.116 8
翻转	1.942 8	0.000 2	0.036 4	0.113 8
缩放	0.459 3	0	0.005 8	0.020 9
高斯模糊	1.737 5	0	0.014 4	0.055 2
锐化	0.500 4	0	0.002 8	0.023 1
滤波	0.082 3	0	0.000 5	0.002 3
像素化	0.084 6	0	0.001 4	0.005 6
高斯噪声	0.343 9	0.000 2	0.010 7	0.023 0
亮度调整	3.060 8	0.000 2	0.027 3	0.103 8
JPEG 压缩	3.034 3	0	0.010 6	0.093 0
色彩空间调整	0.051 5	0	0.000 8	0.002 2
边框叠加	0.276 6	0.000 1	0.007 0	0.016 4
组合操作	4.067 8	0.000 3	0.050 7	0.231 0

Table 3 Accuracy of Perceptual Identical and Distinct Images under Different Thresholds T

表 3 不同阈值 T 下感知相似和不相似图像的正确率

阈值 T	Γ_{sim} /%	Γ_{dif} /%
0.7	95.80	98.19
0.8	96.41	97.64
0.9	**97.14**	**97.10**
1.0	97.51	96.59
1.1	98.00	96.09
1.2	98.20	95.61
1.3	98.61	95.16
1.4	99.10	94.71
1.5	99.10	94.28
1.6	99.31	93.85
1.7	99.31	93.44

注：黑体数值表示最优值.

我们将本文方案与 4 个典型的 PAH 方案进行鲁棒性和唯一性 2 方面的综合比较. 这 4 个方案分别是 RPIVD[28], RE[29], DCP[30], MCCNN[20], 其中前 3 个是基于传统图像处理的 PAH 方案, 最后 1 个是目前为数不

多的基于深度学习的 PAH 方案. 在实验对比中, 将这 4 个方案调整到最佳的内容认证性能, 其中 RPIVD, RE, DCP, MCCNN 生成的哈希码长度分别为 40, 64, 72, 50. 结果如图 7 所示.

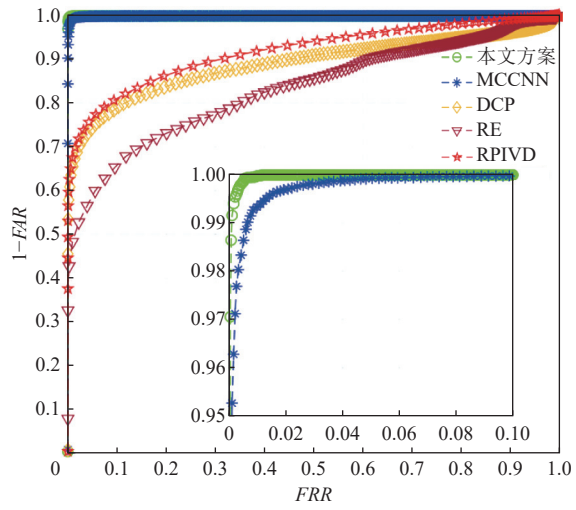


Fig. 7 Comparison of content authentication performance of different PAH schemes

图 7 不同 PAH 方案的内容认证性能比较

图 7 展示了这 5 种 PAH 方案在包含了 31 000 对感知相似图像和 31 000 对感知不相似图像的测试集上的 ROC 曲线. 由图 7 可以看出, 本文方案的 ROC 曲线相比于其他 4 个 PAH 方案更接近于左上角, 表明它具有更好的内容认证性能. 此外, 还利用式(11)计算了本文深度 PAH 方案、RPIVD、RE、DCP、MCCNN 的 $F1$ 分数, 分别为 0.997 1, 0.768 0, 0.841 3, 0.853 7, 0.992 2. 相比而言, 本文方案的 $F1$ 分数值更接近于 1, 进一步说明了本文方案内容认证性能的优越性.

理想的 PAH 方案不仅需要具有出色的内容认证性能, 还需要能够在较短的时间内为图像生成相应的图像哈希码, 即 PAH 方案的哈希生成效率. 为了更好地展示本文方案的哈希生成效率, 与 RE, RPIVD, DCP, MCCNN 还进行了哈希生成效率的比对. 我们从实验数据集中选取了 150 000 张图像来进行测试, 通过计算这 5 个 PAH 方案生成哈希码的平均时间来比较它们的哈希生成效率. 表 4 展示了这 5 个方案哈希生成的平均时间, 可以看出 3 个传统的 PAH 方案的哈希平均生成时间都在几十至 200 ms 之间, 而本文方案和 MCCNN 可以在小于 0.5 ms 的时间内生成图像的哈希码. 这表明, 本文方案有着很高的哈希生成效率, 可以快速地为海量图像数据生成哈希值并进行内容认证.

3.6 消融实验

本文的深度 PAH 模型利用了分类损失与局部和全局特征融合这 2 个策略来优化图像感知哈希的内容认证性能. 为了验证这 2 种策略的有效性, 我们构建了 2 个 PAH 模型, 即哈希损失模型和全局特征模型, 并分别在训练集和测试集上进行训练和测试, 这 2 个模型都是在本文提出的深度 PAH 模型基础上修改得到的. 其中哈希损失模型是将本文模型在训练时去掉分类损失, 只保留基础的哈希损失; 全局特征模型是舍弃局部特征与特征融合步骤, 只使用全局特征来生成哈希码.

Table 4 Time of Generating Hash Sequence for Different PAH Schemes

表 4 不同 PAH 方案生成哈希序列的时间 ms

方案	平均时间
本文方案	0.45
RE	43.82
RPIVD	143.84
DCP	181.63
MCCNN	0.43

图 8 展示了本文模型、哈希损失模型和全局特征模型的内容认证性能的 ROC 曲线, 并且它们的 $F1$ 分数分别为 0.997 1, 0.996 3, 0.995 6. 可以看出, 分类损失与局部和全局特征融合这 2 个策略可以使模型的内容认证性能获得有效的提升.

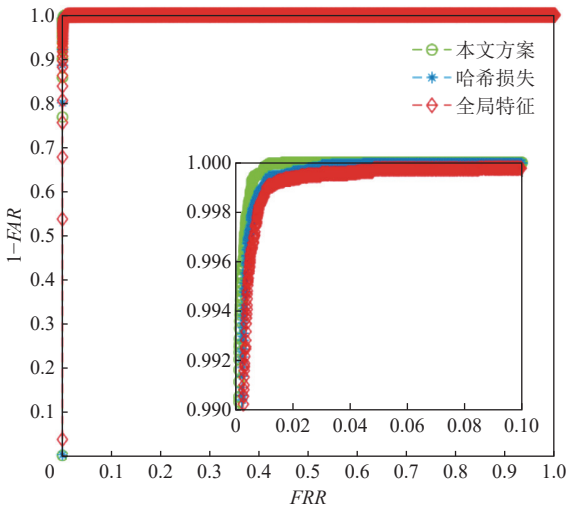


Fig. 8 Content authentication performance under different strategies

图 8 不同策略下的内容认证性能

4 结 论

本文提出了一种基于窗口自注意力机制特征融

合的深度感知认证哈希方案. 运用窗口自注意力机制、局部和全局特征融合、联合损失函数对模型进行优化, 训练出了一个可以有效抵抗图像内容保留操作的感知哈希模型. 其中窗口自注意力机制用于局部特征的生成, 通过将输入的浅层特征按窗口进行分割并计算热力图得到窗口特征, 然后利用其与全局特征之间的相关性选取出最终的局部特征; 特征融合模块用于将局部特征与全局特征进行融合压缩以便后续生成哈希码; 联合损失函数由哈希损失和分类损失构成, 用于约束模型的训练, 增强鲁棒性和唯一性. 大量的实验结果表明, 本文方案的 ROC 曲线和 $F1$ 分数优于已有的代表性 PAH 方案, 说明了本文方案具有较好的图像内容认证性能.

目前本文方案可抵抗的内容保留操作都是已知的, 对于未知或模型未训练过的内容保留操作的鲁棒性现阶段还不够理想. 然而在实际应用中, 图像可能会遭受一些事先未知的操作, 如何实现可有效抵抗未知内容保留操作的 PAH 方案还需进一步深入研究.

作者贡献声明: 周元鼎提出了算法思路和实验方案; 高国鹏和房耀东负责完成实验并撰写论文; 秦川提出指导意见并修改论文.

参 考 文 献

[1] Yan Caiping, Pun Chiman, Yuan Xiaochen. Multi-scale image hashing using adaptive local feature extraction for robust tampering detection[J]. Signal Processing, 2016, 121(C): 1−16

[2] Su Qingtang, Liu Decheng, Sun Yehan. A robust adaptive blind color image watermarking for resisting geometric attacks[J]. Information Sciences: An International Journal, 2022, 606: 194−212

[3] Zhu Xinshan, Chen Yanming, Dong Honghui, et al. Robust double domain watermarking for binary document image[J]. Chinese Journal of Computers, 2014, 37(6): 1352−1364 (in Chinese)
(朱新山, 陈砚鸣, 董宏辉, 等. 基于双域信息融合的鲁棒二值文本图像水印[J]. 计算机学报, 2014, 37(6): 1352−1364)

[4] Liao Xin, Li Kaide, Zhu Xinshan, et al. Robust detection of image operator chain with two-stream convolutional neural network[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(5): 955−968

[5] Peng Anjie, Kang Xiangui. Median filtering forensics based on multi-directional difference of filtering residuals[J]. Chinese Journal of Computers, 2016, 39((3):): 503−515 (in Chinese)
(彭安杰, 康显桂. 基于滤波残差多方向差分的中值滤波取证技术[J]. 计算机学报, 2016, 39(3): 503−515)

[6] Shen Qi, Zhao Yan. Perceptual hashing for color image based on color opponent component and quadtree structure[J]. Signal Processing, 2020, 166: 107244

[7] Zhang Qiuyu, Han Jitian. A novel color image encryption algorithm based on image hashing, 6D hyperchaotic and DNA coding[J]. Multimedia Tools and Applications, 2021, 80(9): 13841−13864

[8] Qin Chuan, Guo Mengqi, Li Xinran, et al. Robust image hashing in encrypted domain[J]. Journal of Software, 2023, 34(2): 868−883 (in Chinese)
(秦川, 郭梦琦, 李欣然, 等. 一种加密域鲁棒图像哈希算法[J]. 软件学报, 2023, 34(2): 868−883)

[9] Huang Ziqing, Liu Shiguang. Perceptual image hashing with texture and invariant vector distance for copy detection[J]. IEEE Transactions on Multimedia, 2020, 23: 1516−1529

[10] Huang Ziqing, Tang Zhenjun, Zhang Xianquan, et al. Perceptual image hashing with locality preserving projection for copy detection[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(1): 463−477

[11] Du Ling, Ho A T S, Cong Runmin. Perceptual hashing for image authentication: A survey[J]. Signal Processing: Image Communication, 2020, 81: 115713

[12] Li Xinran, Qin Chuan, Wang Zichi, et al. Unified performance evaluation method for perceptual image hashing[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1404−1419

[13] Li Xiaoqing, Yang Jiansheng, Ma Jinwen. Recent developments of content-based image retrieval (CBIR)[J]. Neurocomputing, 2021, 452: 675−689

[14] Ouyang Junlin, Liu Yizhi, Shu Huazhong. Robust hashing for image authentication using SIFT feature and quaternion Zernike moments[J]. Multimedia Tools and Applications, 2017, 76(2): 2609−2626

[15] Singh S P, Bhatnagar G, Singh A K. A new robust reference image hashing system[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(4): 2211−2225

[16] Srivastava M, Siddiqui J, Ali M A. Robust image hashing based on statistical features for copy detection[C]//Proc of the IEEE Uttar Pradesh Section Int Conf on Electrical, Computer and Electronics Engineering (UPCON'16). Piscataway, NJ: IEEE, 2016: 490−495

[17] Qin Chuan, Chang C C, Tsou P L. Robust image hashing using non-uniform sampling in discrete Fourier domain[J]. Digital Signal Processing, 2013, 23(2): 578−585

[18] Abdullahi S M, Wang Hongxia, Li Tao. Fractal coding-based robust and alignment-free fingerprint image hashing[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 2587−2601

[19] Li Yuenan, Wang Dongdong, Tang Linlin. Robust and secure image fingerprinting learned by neural network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(2): 362−375

[20] Qin Chuan, Liu Enli, Feng Guorui, et al. Perceptual image hashing for content authentication based on convolutional neural network with multiple constraints[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(11): 4523−4537

[21] Zhang Han, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks [C] //Proc of the 36th Int Conf on Machine Learning. New York: ICML, 2019: 7354−7363

[22] Wang Zhihui, Wang Shijie, Zhang Pengbo, et al. Weakly supervised

fine-grained image classification via correlation-guided discriminative learning［C］//Proc of the 27th ACM Int Conf on Multimedia. New York: ACM, 2019: 1851−1860

［23］ Cao Bingyi, Araujo A, Sim J. Unifying deep local and global features for image search［C］//Proc of the 16th European Conf on Computer Vision. Berlin: Springer, 2020: 726−743

［24］ Hjelm R D, Fedorov A, Lavoie-Marchildon S, et al. Learning deep representations by mutual information estimation and maximization［C］//Proc of the 7th Int Conf on Learning Representations. Washington, DC: ICLR, 2019: 9153−9176

［25］ Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Proc of the 10th European Conf on Computer Vision. Berlin: Springer, 2014: 740−755

［26］ Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]//Proc of the 15th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 618−626

［27］ Douze M, Tolias G, Pizzi E, et al. The 2021 image similarity dataset and challenge［J］. arXiv preprint, arXiv: 2106.09672, 2021

［28］ Tang Zhenjun, Zhang Xianquan, Li Xianxian, et al. Robust image hashing with ring partition and invariant vector distance[J]. IEEE Transactions on Information Forensics and Security, 2015, 11(1): 200−214

［29］ Tang Zhenjun, Zhang Xianquan, Huang Liyan, et al. Robust image hashing using ring-based entropies[J]. Signal Processing, 2013, 93(7): 2061−2069

［30］ Qin Chuan, Chen Xueqin, Luo Xiangyang, et al. Perceptual image hashing via dual-cross pattern encoding and salient structure detection[J]. Information Sciences, 2018, 423: 284−302



Zhou Yuanding, born in 1997. PhD candidate. His main research interests include deep perceptual hashing and multimedia information processing.

周元鼎, 1997 年生. 博士研究生. 主要研究方向为深度感知哈希、多媒体信息处理.



Gao Guopeng, born in 1998. Master. His main research interest includes perceptual image hashing.

高国鹏, 1998 年生. 硕士. 主要研究方向为感知图像哈希.



Fang Yaodong, born in 1997. Master. His main research interests include perceptual image hashing and image authentication.

房耀东, 1997 年生. 硕士. 主要研究方向为感知图像哈希、图像认证.



Qin Chuan, born in 1980. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include multimedia information security and AI security.

秦　川, 1980 年生. 博士，教授，博士生导师. CCF 高级会员. 主要研究方向为多媒体信息安全、AI 安全.