

基于端到端哈希生成模型的大容量映射式无载体信息隐藏技术

贾颜泽¹ 张春玉¹

JIA Yanze ZHANG Chunyu

摘要

无载体信息隐藏技术因其良好的抗隐写分析能力而备受关注。现有技术主要分为生成式和映射式两类方法。与前者相比,映射式方法保证了信息的高提取准确率和更强的鲁棒性。然而,映射式方法仍然面临着设计完备码表映射规则的挑战,抵抗几何攻击与非几何攻击鲁棒性的不平衡以及对大规模图库的依赖等问题。针对这些问题,文章提出了一种基于端到端哈希生成模型的大容量映射式无载体信息隐藏方案。利用哈希生成模型生成深度散列序列,同时利用(sine-tent-logistic, STL)混沌系统置乱图像散列序列来映射秘密信息,利用不同的置乱特征将一幅图像映射到多个秘密信息实例,以此实现秘密信息隐藏,只需传输一张载体图像和混沌参数即可,减少对图像数量的要求,避免不完全映射。信息提取时将载体图像输入相同的哈希生成模型中得到深度散列序列,再将混沌参数和图像散列序列输入 STL 混沌系统中得到混沌后的散列序列,就可得到秘密信息。实验结果表明,该方法隐藏容量得到提高,且无需大规模图库的支撑,减少了所需映射图像数量,同时,还能保持对几何攻击和非几何攻击的鲁棒性。

关键词

无载体信息隐藏;映射式;端到端哈希生成;混沌映射

doi: 10.3969/j.issn.1672-9528.2025.11.043

0 引言

随着互联网和云计算的发展,越来越多的数据通过网络传输,避免数据受到非授权访问的两种主要方法是加密技术和信息隐藏技术。信息隐藏技术以其隐蔽通信过程不可被感知的特点备受青睐。传统信息隐藏技术因嵌入秘密信息的需要而对载体图像进行各种修改,这些修改的痕迹成为隐写分析检测的突破点。因此,无载体信息隐藏思想被提出^[1],该技术不是没有载体,而是强调在嵌入秘密信息时不对载体做任何修改,因而具有高安全性。

无载体信息隐藏近几年来成为新的研究热点,其主要包含两大类:构造生成式和搜索映射式^[2]。本文试图重点研究映射式无载体信息隐藏技术,映射式无载体信息隐藏方法于2015年由Zhou等人^[3]提出,首先在网络上搜集大量图片构成数据库,随后生成图像哈希序列和图像做映射索引,找到哈希序列和秘密信息分段后对应的图像作为隐藏图像。此方法需要大量的图像并且映射覆盖率较低。随后,周志立等人^[4]提出了基于词袋(bag of words, BOW)模型构建视

觉关键词和文本关键词的映射关系库进行信息隐藏,但隐藏容量有限。Liu等人^[5]结合DenseNet特征图像检索和离散小波变换(discrete wavelet transform, DWT)序列映射生成鲁棒特征序列。为进一步增强鲁棒性,Zhou等人^[6]使用快速区域卷积神经网络(faster region-based convolutional neural network, Faster-RCNN)检测和识别图像中的物体,得到鲁棒目标对象标签的序列生成映射规则,从而实现无载体信息隐藏。Meng等人^[7]提出散列生成模型的鲁棒性无载体图像隐写算法。采用端到端的CNN模型生成散列序列,降低了构造数据集的难度,提高了映射效率,但仍有秘密信息无法映射完全。Liu等人^[8]利用词袋模型得到图像坐标序列,将秘密信息用坐标序列表示,但传输坐标序列所需的传输负载过大,远超过仅传输秘密信息所需。

针对上述国内外研究现状,本研究旨在提出高效映射方法,解决对大规模图库的依赖和映射率低的问题,并试图有效提升信息隐藏容量。同时,提高算法抵抗几何攻击和非几何攻击的鲁棒性平衡。

1 相关工作

1.1 图像深度哈希散列

本文使用深度跨模态哈希(deep cross-modal hash-

1. 西藏民族大学信息工程学院 陕西咸阳 712000

[基金项目] 国家自然科学基金“结合领域知识构造唐卡图像的无载体信息隐藏方法研究”(62262062)

ing, DCMH) 模型^[9]中的卷积神经网络作为网络架构, 并加入 (convolutional block attention module, CBAM) 混合注意力机制^[10]模块。DCMH 模型无缝集成特征学习部分和哈希码学习部分。在学习过程中, 每个部分都可以向其他部分提供反馈。如图 1 所示, 该模型的结构主要由两个模块组成, 即特征提取模块和哈希生成模块。其中, 特征提取模块提取图像深度特征, 哈希生成模块将连续的特征映射成散列序列。

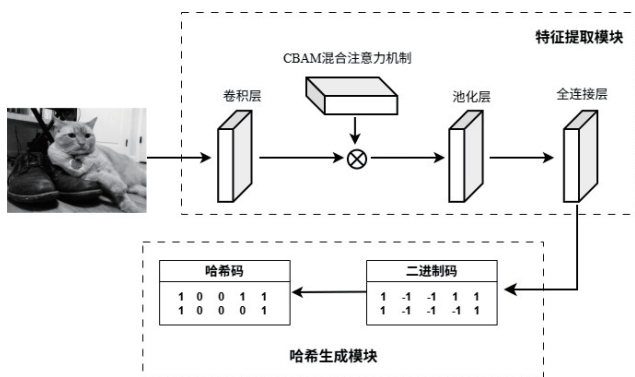


图 1 模型框架图

(1) 特征提取模块

本文以 DCMH 模型中深度卷积网络 CNN 为基准提取图像的深层哈希序列。8 层网络结构包含 5 个卷积层和 3 个全连接层。前 7 层架构与 CNN-F 模型保持一致, 第 8 层设为全连接层, 输出为学习到的图像深度特征, 同时加入 CBAM 混合注意力机制模块提高抵抗攻击的鲁棒性。

(2) 哈希生成模块

令 $f(x_i; \theta_x) \in \mathbb{R}^c$ 表示对于点 i 学习的图像特征, 对应于 CNN 的输出。 θ_x 是 CNN 的网络参数。模型的目标函数定义为:

$$\min_{B^{(x)}, \theta_x} \mathcal{J} = -\sum_{i=1}^n (\theta_i - \log(1 + e^{\theta_i})) + \gamma(\|B^{(x)} - F\|_F^2) + \eta(\|F\|_F^2), B^{(x)} \in \{-1, +1\}^{c \times n} \quad (1)$$

式中: F 表示图像特征。 $F_{*i} = f(x_i; \theta_x)$ 且 $F \in \mathbb{R}^{c \times n}$, $\theta_i = \frac{1}{2} F_i^T F_i$ 。 F 的第 i 列记为 F_{*i} ; $\|\cdot\|_F$ 表示一个矩阵的 Frobenius 范数; $B_{*i}^{(x)}$ 表示图像二进制哈希码; γ 和 η 表示超参数。

优化第二项 $\gamma(\|B^{(x)} - F\|_F^2)$, 可得到 $B^{(x)} = \text{sign}(F)$, 认为 F 是 $B^{(x)}$ 的可连续替代。

$$\text{sign}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (2)$$

第三项 $\eta(\|F\|_F^2)$ 被用作在所有训练点上使哈希序列的每一位均衡。更具体地说, 所有训练点上每个比特的 +1 和 -1 的数量应该几乎相同, 此约束可用于最大化每个位所提供的信息。设置 $B^{(x)} = B$, 式 (1) 问题可转换为最终要学习的目标函数:

$$\min_{B, \theta_x} \mathcal{J} = -\sum_{i=1}^n (\theta_i - \log(1 + e^{\theta_i})) + \gamma(\|B - F\|_F^2) + \eta(\|F\|_F^2), B \in \{-1, +1\}^{c \times n} \quad (3)$$

当 B 固定时, 通过使用反向传播 (BP) 算法来学习图像

模态的 CNN 参数 θ_x 。在大多数现有的深度学习方法中, 利用随机梯度下降 (SGD) 和 BP 算法来进行学习。特别是对于每个采样点 i , 首先计算梯度为:

$$\frac{\partial \mathcal{J}}{\partial F_{*i}} = \sum_{i=1}^n \left(\frac{1}{1 + e^{-\theta_i}} F_{*i} - F_{*i} \right) + 2\gamma(F_{*i} - B_{*i}) + 2\eta F_{*i} \quad (4)$$

然后使用链式法则利用 $\frac{\partial \mathcal{J}}{\partial F_{*i}}$ 计算 $\frac{\partial \mathcal{J}}{\partial \theta_x}$ 。

对于不在训练集中的任何点, 如给定点 q 的图像模态 X_q , 采用前向传播来生成哈希序列为:

$$b_q^{(x)} = h^{(x)}(x_q) = \text{sign}(f(x_q; \theta_x)) \quad (5)$$

前向传播生成哈希序列后, 最终图像对应的二进制哈希序列如式 (6) 所示。若 $b_q = +1$, 则不做处理; 若 $b_q = -1$, 将其变为 0。

$$B = \begin{cases} 1, & b_q = +1 \\ 0, & b_q = -1 \end{cases} \quad (6)$$

1.2 Sine-Tent-Logistic 混沌系统

本文使用的 (sine-tent-logistic, STL) 混沌系统^[11], 由 Sine 映射、Tent 映射和 Logistic 映射构成, 组合后的系统可保持较高的灵敏度。与相应的单种子映射相比, STL 具有更复杂的混沌特性。此外, 当其中一个种子映射超出混沌范围时, 仍能保持良好的混沌行为。3 个映射函数分别为:

$$S_{n+1} = S(\mu_s, S_n) = \mu_s \sin(\pi S_n) \quad (7)$$

$$T_{n+1} = T(\mu_T, T_n) = 1 - \mu_T |T_n - 0.5| \quad (8)$$

$$L_{n+1} = L(\mu_L, L_n) = \mu_L L_n (1 - L_n) \quad (9)$$

式中: $\mu_s \in (0, 1]$, $\mu_T \in (0, 2]$, $\mu_L \in (0, 4]$ 。构造 STL 映射, 将 x_n 先输入到 Logistic 映射和 Tent 映射融合结构中处理并进行相加, 将结果作为 Sine 映射的输入, 最终得到输出 x_{n+1} , 经过 K 次迭代后, 级联结构即可产生长度为 K 的随机序列。

STL 映射表达为:

$$x_{n+1} = \begin{cases} \sin(\pi((1 - 2|x_n - 0.5|) + (\mu_{x_n}(1 - x_n)))) - 1, & x_n > 1 \\ \sin(\pi((1 - 2|x_n - 0.5|) + (\mu_{x_n}(1 - x_n)))) & 0 \leq x_n \leq 1 \\ \sin(\pi((1 - 2|x_n - 0.5|) + (\mu_{x_n}(1 - x_n)))) + 1, & x_n < 0 \end{cases} \quad (10)$$

式中: $4 \geq \mu > 0$

2 算法流程介绍

2.1 信息隐藏算法

具体的信息隐藏算法步骤如下, 框架图如图 2 所示。

第一步: 计算图像特征哈希序列。选取一张载体图像 I , 根据 1.1 中的方法计算基于图像特征的深度哈希序列 B 。

第二步: 秘密信息分段。将长度为 N 的秘密信息 S 分成长度为 n 的 m 个二进制序列段, 如果长度 N 是不可整除的, 则在最后一段中补充 q 个零以保证每个秘密信息二进制序列

都具有 n 比特:

$$m = \begin{cases} N/n, & \text{if } H\%n = 0 \\ \lfloor N/n \rfloor + 1, & \text{otherwise} \end{cases} \quad (11)$$

由上述公式可得, 秘密信息可表示为 $S = [S_{s1}, S_{s2}, \dots, S_k, \dots, S_m]$ 。

第三步: 运行 1.2 节 STL 混沌系统。输入随机数 x , 其范围在 $-1\ 000 \leq x \leq 1\ 000$ 。迭代 K ($K \gg n$) 次, 这里迭代 500 次以确保混沌性^[12], 生成长度为 K 的混沌序列 C , 并截取混沌序列最后 n 个数据, 记为 X :

$$X = [c_{n+1}, c_n, \dots, c_0, \dots, c_0] \quad (12)$$

式中: c_i 表示混沌序列 C 的倒数第 $i+1$ 个值。

第四步: 对 X 进行残差量化。 X 为十进制格式, 与图像特征哈希序列格式不同。为了确保混合混沌序列和图像特征哈希序列具有相同的格式, 得到二进制序列 H 。

$$H(i) = \begin{cases} 1, & X(i) > X(i+1) \\ 0, & X(i) \leq X(i+1) \end{cases}, 0 \leq i \leq n-1 \quad (13)$$

第五步: 得到混沌特征序列。使用混沌序列 H 对图像特征的深度哈希序列 B 进行加扰生成混沌特征序列 E , 其公式为:

$$E = B \oplus H \quad (14)$$

式中: $\oplus$ 是位异或 (XOR) 运算。

第六步: 匹配算法。将混沌特征序列 E 与秘密信息段 S_k 进行匹配, 如果匹配不成功, 则执行第三步; 如果匹配成功, 则记录初始参数, 定义了二元组 (x, μ) 来保存混沌初始参数, 在元组中, x 对应输入的随机初值; μ 为参数。并判断是否匹配完最后一个秘密信息段, 如果没有匹配完, 则执行第三步; 若匹配结束, 执行第七步。

第七步: 传输载体图像 I 及辅助信息。将记录的 m 个混沌系统初始参数按顺序组成序列, 把第二步中补零的个数 q 构造为二元组 $\{0, q\}$ 并添加在二元组序列最后面, 连同载体图像 I 一起传输给接收方。

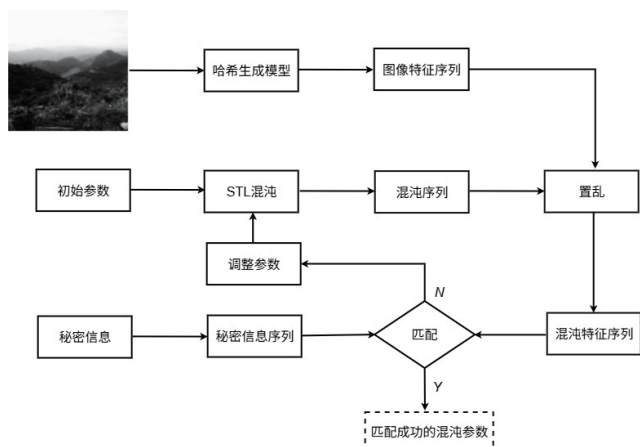


图2 信息隐藏阶段框架图

2.2 信息提取算法

本算法的图像深度哈希散列模型和 STL 混沌映射系统由收发双方共享, 传输过程中只传输一张载体图像和混沌初始参数序列, 如图 3 所示。

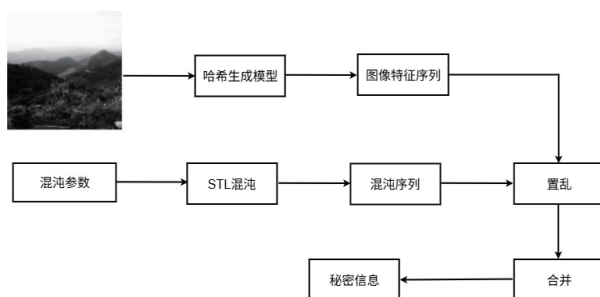


图3 信息提取阶段框架图

具体的信息提取算法步骤如下:

第一步: 将接收到的载体图像输入到图像深度哈希散列模型中, 得到对应的特征哈希二进制序列 B 。

第二步: 将混沌初始参数最后的二元组提取出来, 得到补零的个数 q 。依次将 m 个混沌初始参数二元组按顺序分别输入到 STL 混沌映射系统中, 得到相对应的混沌序列 H_k 。

第三步: 将图像特征哈希序列 B 依次分别和第二步中生成的混沌序列 H_k 做位异或运算来恢复秘密信息段 S_k , 并按顺序组合为秘密二进制序列 S :

$$S_k = B \oplus H_k \quad (15)$$

将秘密信息二进制序列 S 除去尾部的 q 个 '0' 恢复出最终的秘密信息。

3 实验及分析

本方法的 CNN 模型是在一个带有 PyTorch 框架的 NVIDIA GeForce RTX 3090 上训练完成。其他实验部分在 Intel(R) Core(TM) i5-1235U 1.30 GHz、16 GB 存储器的个人计算机上完成。本文实验利用 MS COCO 2014 训练集对 CNN 模型进行训练且从 MS COCO 2014 数据集中随机选择了 500 张图像进行实验测试。本文算法与文献 [5-7] 在 MS COCO 2014 数据集上进行鲁棒性比较实验。

对比实验中的参数设置为: 哈希序列的长度设置为 8; 所有图像的大小都被调整为 224×224 以进行实验; 超参数 γ 和 η 均为 1; batch_size 为 32; 训练轮数为 100 轮; 学习率为 $10^{-1.5}$ 。

3.1 鲁棒性分析

在传输过程中, 图像不可避免地会受到很多攻击的破坏, 即几何攻击和非几何攻击。本算法具有较强的鲁棒性, 本实验中鲁棒性是通过受攻击后提取精度来衡量的, 提取精度越高, 鲁棒性越强。提取精度 ACC 定义为:

$$ACC = \sum_{i=1}^{\text{len}} \varphi(bs_i, bs'_i) / \text{len} \times 100\% \quad (16)$$

散列序列长度为 len , $\sum_{i=1}^{len} \varphi$ 用于统计两个序列对应位相同的个数。 φ 定义为:

$$\varphi(bs_i, bs'_i) = \begin{cases} 1, bs_i = bs'_i \\ 0, bs_i \neq bs'_i \end{cases} \quad (17)$$

图像生成的特征哈希序列表示为 bs_i ; 被攻击图像中提取的特征哈希序列表示为 bs'_i 。

由表 1 数据可得, 本方法在平衡抵抗几何攻击与非几何攻击鲁棒性上整体优于其他方法。人为设计、固定的特征提取方式和哈希生成方式在面对不同类型的攻击时无法进行调整, 本方法使用 DCMH 模型中的卷积神经网络作为主干网络, 加入 CBAM 混合注意力机制进一步提高鲁棒性; 同时 DCMH 模型直接学习离散哈希码, 提高了学习哈希码的准确性。文献 [7] 方法虽与本文方法都使用端到端哈希生成模型, 但本文方法平衡抵抗几何攻击与非几何攻击鲁棒性能好于文献 [7]。

表 1 MS COCO2014 数据集上鲁棒性比较

攻击类别	参数	各种算法的鲁棒性 /%			
		文献 [5]	文献 [6]	文献 [7]	本文算法
旋转	10°	9.7	55.4	92.2	94.5
	30°	4.3	33.8	81.2	93.2
	50°	3.2	22.9	67.8	91.1
边缘裁剪	5°	66.4	74.5	98.2	95.9
	10°	59.0	72.1	98.2	95.2
	20°	54.3	64.3	98.2	91.7
中心裁剪	5°	84.6	56.3	90.0	99.0
	10°	76.8	32.0	84.8	97.5
	20°	74.4	23.4	70.0	96.5
平移	(36.20)	7.1	65.6	93.4	94.7
	(40.25)	6.5	66.2	93.2	93.9
	(80.50)	3.4	56.1	91.6	94.0
中值滤波	{3×3}	92.7	60.4	95.6	98.9
均值滤波	{3×3}	96.6	63.0	95.0	98.8
高斯滤波	{3×3}	97.0	64.5	95.6	98.9
JPEG 压缩	$Q=90$	97.4	76.4	98.2	99.2
椒盐噪声	$\sigma=0.001$	95.1	71.0	97.0	98.7
高斯噪声	$\sigma=0.001$	93.2	65.4	97.0	95.5
伽马变换	0.8	92.2	79.2	97.4	96.7
平均		58.6	58.0	91.3	96.0

3.2 容量分析

映射式无载体信息隐藏方法中, 图像映射编码的长度越长, 每张图像的隐藏容量越大, 需要传输的载密图像越少, 就可以保证更高的系统安全性。然而, 这时需要设计的完备码表的映射规则就越难, 往往图像映射编码的长度增加会导

致图库中所需候选图像数量呈指数增长。因此, 隐藏容量不足成为掣肘映射式无载体信息隐藏方法发展的主要瓶颈, 考虑隐藏容量时也应关注候选图像数量, 本节用比率 R_{mci} 对本文算法与相关算法的单张图像有效隐藏容量进行公平、直观地对比分析。定义为:

$$R_{mci} = M/N_{ci} \quad (18)$$

式中: R_{mci} 表示每张图像有效的隐藏容量; M 表示隐藏容量; N_{ci} 表示隐藏 M 位秘密信息所需载体图像数量。 R_{mci} 越大, 表示单位图像的隐藏容量越大。本算法使用的 DCMH 神经网络目前可生成的图像特征序列最大长度为 64, 本实验中将 M 设为 64, 实验结果列于表 2 中。

表 2 单位图像有效隐藏容量对比表

方法类别	M	N_{ci}	R_{mci}
文献 [5]	15	2^{15}	4.58×10^{-4}
文献 [6]	6	2^6	9.38×10^{-2}
文献 [7]	15	2^{15}	4.58×10^{-4}
本文算法	64	1	6.4×10^1

从表 2 中可以看到, 本方法的 R_{mci} 值为 64, 明显优于另外 4 种方案。其根本核心在于本方案采用 STL 混沌系统, 通过灵活选择不同的混沌参数使得单张候选图像可以映射到多个秘密信息实例, 大幅降低了对候选图像数量的需求, 使得单张候选图像的有效隐藏容量性能得到有效提升。虽然本文方法需要传递额外的混沌参数, 但是其相对于其他方案对图库指数级的依赖和大量载体图像的传输负载, 本文不依赖大规模图库, 传输负载也大大降低了。而文献 [5-7] 的方法都需要建立图像库。

4 结论

本文提出了一种基于端到端哈希生成模型的大容量映射式无载体信息隐藏方法。利用哈希生成模型直接获得图像深度哈希序列, 同时结合 STL 混沌映射方法实现信息隐藏, 无需建立大规模图像库和索引表, 只需传输一张载体图像和混沌映射参数。本方法增加了信息隐藏容量, 提高了平衡抵抗几何攻击和非几何攻击的鲁棒性。

参考文献:

[1]SHI Y Q, CHEN C H, CHEN W.A markov process based approach to effective attacking JPEG steganography [C]//Information Hiding. Berlin: Springer, 2006: 249-264.

[2]QIN J H, LUO Y J, XIANG X Y, et al. Coverless image steganography:a survey[J].IEEE access,2019,7:171372-171394.

[3]ZHOU Z L, SUN H Y, HARIT R,et al.Coverless image steganography without embedding[C]//Cloud Computing and Security.Berlin:Springer, 2015:123-132.

基于生成式强化学习的代码缺陷修复生成方法

罗莉琼¹

LUO Liqiong

摘要

代码缺陷修复作为软件工程中的关键环节,其自动化智能化程度直接影响系统的健壮性与开发效率。文章研究了一种基于生成式强化学习的代码缺陷修复方法,构建缺陷上下文的状态空间与修复动作空间,引入多维特征组合设计回报函数,建立融合预训练模型与DDPG策略优化的修复生成架构,分析不同回报设计对训练收敛性的影响,探讨模型在真实缺陷数据集上的修复性能与泛化能力表现。该方法有助于提升代码修复过程中的智能决策水平,为面向复杂代码环境的自适应缺陷修复提供技术路径。

关键词

生成式强化学习;代码缺陷修复;深度强化学习

doi: 10.3969/j.issn.1672-9528.2025.11.044

0 引言

强化学习(reinforcement learning, RL)是一种机器学习方法,强化学习的基础框架是马尔可夫决策过程,允许智能体(Agent)能够在与环境(Environment)的交互中通过试错来学习最优策略。智能体在环境中执行行动(Action),并根据行动的结果接收反馈,即奖励(Reward)。这些奖励信号指导智能体调整其策略,以最大化长期累积奖励。软件

开发过程中的代码缺陷频繁出现,传统修复方法依赖人工经验,效率低且难以适应复杂语境。生成式强化学习具备自动探索策略与动态优化的能力,为代码缺陷修复提供了新思路。该方法可将缺陷定位与修复过程建模为马尔可夫决策问题,以状态-动作序列驱动修复建议生成,结合回报函数引导模型学习有效策略。构建合理的回报函数对修复效果有显著影响,策略优化机制直接决定修复生成质量,强化学习框架的选择影响训练效率与泛化能力。围绕问题建模、策略设计与实验验证三个层面展开,将为实现高质量的自动代码修复奠

1. 株洲开放大学 湖南株洲 412000

- [4] 周志立,曹斌,孙星明.基于图像 Bag-of-Words 模型的无载体信息隐藏[J].应用科学学报,2016,34(5):527-536.
- [5] LIU Q, XIANG X Y, QIN J H, et al. Coverless steganography based in image retrieval of densenet features and DWT sequence mapping[J]. Knowledge-based systems, 2020, 192:105375.
- [6] ZHOU Z L, CAO Y, WANG M M, et al. Faster-RCNN based robust coverless information hiding system in cloud environment[J]. IEEE access, 2019, 7:179891-179897.
- [7] MENG L J, JIANG X H, ZHANG Z Z, et al. A robust coverless image steganography based on an end-to-end hash generation model[J]. IEEE transactions on circuits and systems for video technology, 2023, 33(7): 3542-3558.
- [8] LIU H L, ZHANG C Y, WANG Z J, et al. To deliver more information in coverless information hiding[J]. Multimedia tools and applications, 2023, 83:7215-7229.
- [9] JIANG Q Y, LI W J. Deep cross-modal hashing[C]//2017

IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE, 2017: 3270-3278.

- [10] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Berlin: Springer, 2018: 3-19.
- [11] 陈云攀. 基于小波变换和遗传算法的混沌图像加密系统研究[D]. 南昌:江西财经大学, 2023.
- [12] 赵欣. 不同一维混沌映射的优化性能比较研究[J]. 计算机应用研究, 2012, 29(3): 913-915.

【作者简介】

贾颜泽(2000—),女,山东枣庄人,硕士研究生,研究方向:信息隐藏和数字图像处理。

张春玉(1979—),女,陕西汉中,人,硕士,硕士生导师、副教授,研究方向:信息隐藏与数字水印、数字图像处理。

(收稿日期:2025-04-11 修回日期:2025-10-30)