

基于预处理层增强和注意力机制的 空域图像隐写分析

罗维薇, 刘长龙, 雷 琴

(兰州交通大学 电子与信息工程学院, 兰州 730070)

摘 要: 为了更全面地捕获隐写操作对图像统计特征的改变, 提高空域隐写分析的检测精度, 本文结合隐写算法的嵌入特点, 采用导数和 Gabor 双重滤波器对图像进行预处理, 并对滤波器提取进行增强, 产生多种残差图像, 有效增加隐写分析特征的多样性。将优化的 CBAM 模块嵌入残差块中, 引导网络有效聚焦于具有丰富隐写信号的区域, 从而提高网络的判别学习能力和训练效果。将本模型与经典模型在 BOSSbase v1.01 和 BOWS2 两个公开数据集上进行比较, 实验结果表明: 该方法的检测精度优于现有主流模型 Ye-Net、SRNet 和 ZhuNet。

关键词: 隐写分析; 卷积神经网络; 预处理层增强; 注意力模块

中图分类号: TP309 **文献标志码:** A **文章编号:** 1671-5497(2025)12-4024-10

DOI: 10.13229/j.cnki.jdxbgxb.20240437

Spatial image steganography based on preprocessing layer enhancement and attention mechanism

LUO Wei-wei, LIU Chang-long, LEI Qin

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: In order to capture the changes of statistical features of images caused by steganography more comprehensively and improve the detection accuracy of spatial steganalysis, the embedding characteristics of the steganographic algorithm are incorporated. The derivative and Gabor double filter is used to preprocess the image, and the filter extraction is enhanced to produce a variety of residual images, which effectively increases the diversity of steganographic features. The optimized CBAM module is embedded into the residual block to guide the network to effectively focus on the region with rich steganographic signals, thus strengthening the discriminant learning ability and training effect of the network. The proposed model is compared with the classical model on BOSSbase v1.01 and BOWS2, and the experimental results show that the detection accuracy of the proposed method is superior to the existing mainstream models of Ye-Net, SRNet and ZhuNet.

Key words: steganographic analysis; convolutional neural network; pretreatment layer enhancement; attention module

收稿日期: 2024-04-23.

基金项目: 国家自然科学基金项目(62362047).

作者简介: 罗维薇(1977-), 女, 副教授, 硕士. 研究方向: 图像处理, 模式识别, 人工智能.

E-mail: luoweivei@lztu.edu.cn

通信作者: 雷琴(1980-), 女, 副教授, 硕士. 研究方向: 人工智能, 数据挖掘. E-mail: leiqlin@mail.lztu.cn

0 引言

隐写术是信息隐藏的一种主要技术^[1],它通过将秘密信息嵌入图像、视频、音频等数字载体中,从而在不引起第三方怀疑的情况下进行隐蔽通信。由于隐蔽通信的过程很难被第三方察觉和检测,具有很强的欺骗性,不法分子可利用隐写术来危害社会稳定和国家安全。作为隐写术的对抗技术,隐写分析致力于检测载体中是否隐藏了秘密信息,从而发现并阻止基于隐写的非法通信行为。因此,为了有效遏制隐写术的滥用,近年来,隐写分析受到学术界越来越多的关注。

空域隐写分析是指通过提取图像的空域统计特征来判断图像是否加密^[2]。尽管隐写操作对图像视觉特性影响不大,但它会使图像的一些统计特征产生微妙变化。因此,隐写分析的关键在于如何挖掘出隐写操作前后具有显著差异的图像统计特征。早期,研究者通过使用统计矩、特征函数中心等统计信息和简单的分类器来检测图像是否包含了秘密信息,而近年来逐渐发展为通过使用多个相邻像素之间的相关性或频域系数设计高维的特征提取方法来捕获与隐写术相关的复杂统计特征,代表方法有 SRM^[3]、DCTR^[4]和 PHARM^[5]。尽管这些方法已经取得了较好的性能,但随着隐写术的不断发展,有效的高维统计特征提取方法越来越难以设计,并且特征参数的调整耗时且成本很高,这限制了隐写分析的发展。

近年来,随着深度学习的兴起,研究人员发现,神经网络可以学习图像的多层次特征,从而减少了启发式的特征设计,并能更好地反映图像的本质特征^[2]。在众多的深度学习模型中,卷积神经网络(CNN)^[6]因其特有的网络属性而备受关注。相比于传统的依赖手工设计特征的隐写分析方法,CNN能够更好地表示图像的深层信息,并且CNN还可以将特征提取和分类模块集成在一个结构中,这在一定程度上降低了人工成本并提高了模型的性能。因此,基于CNN的隐写分析逐渐成为图像隐写分析的主流方法。

为了提高模型的检测准确性,当前的隐写分析模型引入了各种技术。在网络结构方面,使用了一些经典的卷积神经网络结构,如深度可分离卷积^[7]、残差连接^[8]和 Inception^[9]。在学习策略方面,引入了迁移学习^[10]和数据增强^[11]。然而,现有的网络模型仍存在一些不足,为了获取丰富的

残差图像来提高模型的性能,研究人员仅仅是通过旋转等操作增加高通滤波器的数量,却没有为不同方向上的滤波器建立实际的联接,导致每个方向上的残差是分散的,这对隐写特征提取是不利的,影响了隐写分析的检测性能;另一方面,目前提出的隐写分析模型在训练过程中缺乏对图像纹理复杂区域特征的关注,网络的表达能力受到限制,模型的检测精度也受到了影响。

本文的主要目标是通过使用多类别高通滤波器增强隐写特征的提取能力,并有效区分隐藏在复杂纹理区域中的隐写信息。提出了一种基于预处理层增强和注意力机制的网络模型,该模型的检测精度和泛化能力都表现良好。本文主要贡献如下:

(1) 提出了一种双重滤波器的预处理层增强方法,采用导数和 Gabor 双重滤波器对图像进行预处理,并使用平均值融合方法对滤波器提取进行增强,为模型产生多种残差图像,使得后续卷积层可以提取更多的隐写特征。

(2) 引入了一种高效通道注意力模块 ECA (Efficient channel attention) 对原始的 CBAM (Convolutional block attention module) 模块进行了优化,并将其应用到隐写分析模型中,验证了改进后的注意力模块对模型的检测能力是有效的。

1 相关工作

随着深度学习技术的发展,许多研究人员将深度学习应用到隐写分析领域中,并取得了良好检测效果^[12]。一种有效提高空域隐写检测性能的方法是在图像进入卷积网络之前用高通滤波器对图像进行预处理操作,其目的是滤除大量不相关的图像内容信息,以便随后的卷积层更容易提取图像中的隐写特征。Qian 等^[13]提出的隐写分析网络 GNCNN,使用了一个 5×5 固定的高通滤波器对图像进行预处理,使模型可以直接从残差图像中学习,减少了图像本身内容对特征提取时的干扰。Xu 等提出了 Xu-Net^[14],该网络在框架上仍然沿用 GNCNN 的网络架构特点,并在网络的前端添加了一个固定的高通滤波器对图像进行预处理。同时,该网络还引入了 ABS (Absolute layer) 层、Tanh 激活函数、批归一化 (Batch normalization, BN) 操作和 1×1 卷积,使得网络的性能与 SRM 具有相似的检测效果,甚至优于 SRM 方法。Ye 等^[15]提出的 Ye-Net,利用 SRM 模型中

的 30 个高通滤波器作为预处理层,允许模型学习更多的特征。为了更好地适应隐写噪声分布,提出了一种新型的截断(Truncated linear unit, TLU)激活函数,使得经过卷积后的特征图具有更好的区分性。实验表明:Ye-Net 的性能优于 SRM 方法。ZhuNet^[16]模型改进了 SRM 的 30 个卷积核来提取噪声残差,并使用深度可分离卷积来捕获残差的信道关系。此外,该网络还引入了空间金字塔池化来代替全局平均池化,可以更好地提取局部特征,提高特征表达能力,并允许任意图像大小输入。SNMC-Net^[17]使用了 SRM 滤波核和随机初始化的双通道预处理模块。同时,该网络引入了非局部运算和多通道卷积 NMC 块,以便捕捉不同区域中的高频和低频特征,网络中还添加了 SPP 结构,用来降低特征图维度的同时便于提取更多种类的多阶隐写特征。

目前,基于卷积神经网络的图像隐写分析研究重点大多集中在模型结构的优化上,而在预处理中对高通滤波器的使用仍局限于滤波器数量的简单调整,导致残差图像单一,这并不能达到最大

化提升模型性能的目的。此外,现有的隐写分析模型大多数不加区分地学习并提取图像中的特征,这种情况可能使模型在训练时无法关注到真正包含隐写信息的区域,导致隐写检测精度下降。本文的工作准确地关注到这些问题,并对其进行有效的解决。

2 本文算法

本文提出了一种结合双重滤波器增强和优化视觉注意力模块的深度残差网络。在本节中,首先介绍整体的网络结构,然后分别描述网络中组件设计与改进的动机,并通过实验验证了其有效性。

2.1 网络架构

本文提出的总体网络结构如图 1 所示,其中所有卷积层的核大小都为 3×3 ,非线性激活函数为 ReLu。它接收大小为 256×256 的灰度图像作为输入,并输出一个二分类判定,即该图像是否为载密图像。该网络由 3 个部分组成:①预处理阶段。其目标主要是通过双重滤波器获取多类别噪声残差图,便于随后隐写特征的提取。②特征提

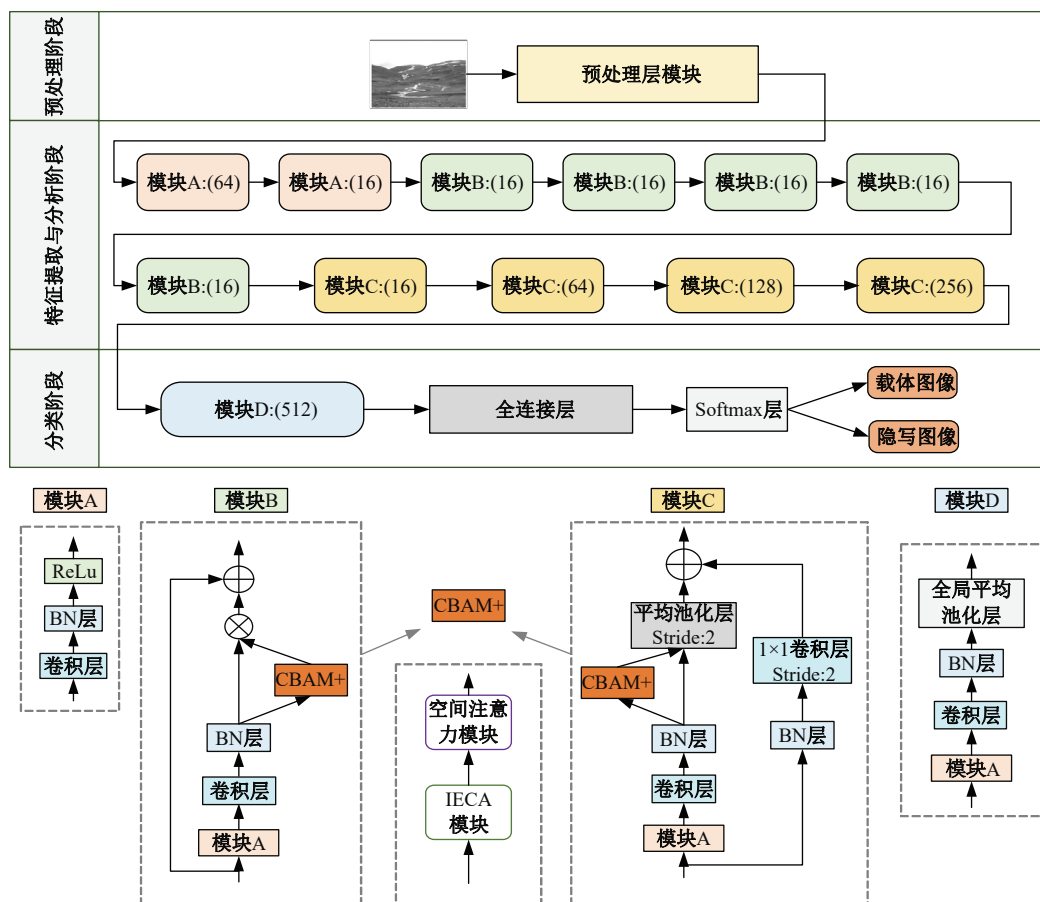


图 1 网络总体结构图

Fig. 1 Overall network structure

取与分析阶段。它负责提取输入图像的隐写噪声信号,并降低特征图的维度。③分类阶段。根据所提取的特征信息来判断图像是否包含秘密信息。

隐写操作可以视作向图像添加极低幅度的噪声信号,隐写噪声的信噪比极低,若不对其进行处理直接输入卷积层,难以提取到有效特征。预处理操作可以有效减少图像本身的内容,放大隐写信号,提高特征提取的效率。因此,在预处理阶段,本文采用导数和 Gabor 双重滤波器对输入图像进行预处理,并对滤波器提取进行了增强,产生了 32 个残差图像,2.2 节将对预处理层增强进行详细描述。

噪声提取与分析阶段由模块 A、模块 B 和模块 C 三种类型的卷积块组成。网络的第 1~7 层从噪声信息中提取统计特征以生成特征图,包含两种类型的卷积块,其中模块 A 是由 1 个卷积层、1 个 BN 层和 ReLU 激活函数构成;模块 B 是在模块 A 基础上加了 1 个卷积层、1 个 BN 层和一个 CBAM+ 模块,其中 CBAM+ 模块是对原始的 CBAM 模块进行了优化,将原始的通道注意力模块进行了替换以避免模型太复杂和出现降维的问题,其优化原理将在 2.3 节进行具体描述,而残差连接是为了避免网络出现梯度消失。由于池化操作不可避免地会损失隐写信号,因此,在网络的前 7 层并未使用池化层。第 8~11 层为特征分析模块,它的主要构成如图 1 中模块 C 所示,其中利用平均池化来实现降低特征图的维数,使用步长为 2 的 1×1 卷积来确保跳跃连接输出相同维度,经过特征分析模块后输出 256 通道数大小为 16×16 的特征图。

分类阶段模型采用全局平均池化集成特征,然后通过全连接层和 Softmax 层输出判别结果。

2.2 预处理层

在隐写分析中,隐写分析特征往往是在残差图像上构建的。因此,丰富的残差图像对隐写特征提取是至关重要的。在本文中,结合了隐写算法擅长在图像纹理复杂区域隐写并改变相邻像素间的相关性的特点,通过利用导数滤波器和 Gabor 滤波器共同作用于预处理层,为模型提供多种类残差图。以下将对这两种滤波器进行详细描述,并通过实验验证对滤波器提取增强的有效性。

(1) 导数滤波器。Li 等^[18]提出的一种基于

多阈值局部二值模式和共生矩阵的隐写分析特征提取方法,其中使用了一组用于隐写分析的导数滤波器。导数通常用来测量信号的瞬时变化率,导数滤波器具有明确的数学意义,它不仅可以作为高通滤波器来抑制图像内容,还可以有效度量像素之间复杂的相关性以放大隐写噪声信号。

对于一幅数字图像 $f(x, y)$, x 和 y 表示图像空间坐标,它在水平方向的一阶偏导和 n 阶偏导可以分别用式(1)和式(2)表示:

$$\frac{\partial f(x, y)}{\partial x} = f(x, y) - f(x + 1, y) \quad (1)$$

$$\frac{\partial^n f(x, y)}{\partial x^n} = \frac{\partial^{n-1} f(x, y)}{\partial x^{n-1}} - \frac{\partial^{n-1} f(x, y)}{\partial (x + 1)^{n-1}} \quad (2)$$

类似地, $f(x, y)$ 沿水平方向的 n 阶和竖直方向的 m 阶混合偏导表达式可以定义如下式所示:

$$\begin{aligned} \frac{\partial^m}{\partial y^m} \left(\frac{\partial^n f(x, y)}{\partial x^n} \right) = & \frac{\partial^{m-1}}{\partial y^{m-1}} \left(\frac{\partial^n f(x, y)}{\partial x^n} \right) - \frac{\partial^{m-1}}{\partial (y + 1)^{m-1}} \left(\frac{\partial^n f(x, y)}{\partial x^n} \right) \end{aligned} \quad (3)$$

对离散信号求偏导等价于采用导数滤波器进行滤波。设 $D_{n,m}$ 表示沿水平方向和垂直方向的阶数分别为 n 阶和 m 阶的导数滤波器,由于混合偏导是可分离的,因此, $D_{n,m}$ 可以通过式(4)中 $D_{n,0}$ 与 $D_{0,m}$ 卷积运算得到。

$$D_{n,m} = D_{n,0} \otimes D_{0,m} \quad (4)$$

例如:

$$\begin{aligned} D_{2,1} = D_{2,0} \otimes D_{0,1} = & \begin{bmatrix} -1 & 2 & -1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \\ & \begin{bmatrix} -1 & 2 & -1 \\ 1 & -2 & 1 \end{bmatrix} \end{aligned} \quad (5)$$

高阶的导数滤波器可以提取更大范围内的隐写噪声信号,但像素间的相关性会随着距离的增大而降低。因此,选择的导数滤波器在某一方向上一般不会超过中阶。本文采用了 16 个导数滤波器,可分为 10 类,列举如下: $D_{1,0}$ 、 $D_{2,0}$ 、 $D_{3,0}$ 、 $D_{4,0}$ 、 $D_{2,1}$ 、 $D_{4,2}$ 、 $D_{1,1}$ 、 $D_{2,2}$ 、 $D_{3,3}$ 、 $D_{4,4}$, 如表 1 所示。表 1 中,将第 1 组滤波器旋转 90° 产生一个新的滤波核,而第 2 组滤波核权重是对称矩阵,因此,不需要对其进行旋转且只包含一个滤波器。表 1 中其余类别的滤波器的基本权重、旋转角度和滤波器数量的含义与上述相同。对于基本权重较短的滤波器(例如,类别 1),若用零填充为 5×5 统一尺寸滤波核之后,零的数量要比实际权重数量多,当

滤波器参与网络学习时,在存在大量零的情况下,学习效果就会大大降低。因此,为了避免这些问题,本文根据其基本权重大小执行不同的补零填充策略,将长度小于等于3的滤波器统一填充至 3×3 大小,长度大于3的滤波器统一填充至 5×5 大小尺寸。不同尺度的卷积核可以捕捉到图像中不同大小的特征,可以更全面地感知图像中的特征信息,从而提高特征提取的准确性。

表 1 滤波器详细信息

Table 1 Filter details

滤波器 类型	基本权重	旋转 角度/(°)	滤波器 数量/个
类别 1	$D_{1,0} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	90	2
类别 2	$D_{1,1} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$	—	1
类别 3	$D_{2,0} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	90	2
类别 4	$D_{2,1} = \begin{bmatrix} -1 & 2 & -1 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$	90	2
类别 5	$D_{3,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -3 & 3 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	90	2
类别 6	$D_{4,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	90	2
类别 7	$D_{4,2} = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ -2 & 8 & -12 & 8 & -2 \\ 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	90	2
类别 8	$D_{2,2} = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$	—	1
类别 9	$D_{3,3} = \begin{bmatrix} 1 & -3 & 3 & -1 \\ -3 & 9 & -9 & 3 \\ 3 & -9 & 9 & -3 \\ -1 & 3 & -3 & 1 \end{bmatrix}$	—	1
类别 10	$D_{4,4} = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ -4 & 16 & -24 & 16 & -4 \\ 6 & -24 & 36 & -24 & 6 \\ -4 & 16 & -24 & 16 & -4 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$	—	1

(2) Gabor 滤波器。Gabor 滤波器是一个可以用于纹理分析的线性滤波器,并且可以在特定方向上分析具有特定频率的图像,因此,把它作为导数滤波器的补充,丰富模型残差图像的类别。Gabor 滤波器可以被定义为高斯函数和余弦函数的乘积,其表达式如下式所示:

$$g(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda}\right) \quad (6)$$

式中: $x' = x \cos \theta + y \sin \theta$, $y' = -x \sin \theta + y \cos \theta$,

在本文实验中滤波器的方向参数 $\theta \in \left\{0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}\right\}$, 尺度参数 $\sigma \in \{0.5, 1\}$, 余弦函数的波长 $\lambda = 0.56\sigma$, 高斯椭圆率的空间纵横比 $\gamma = 0.5$, 因此,产生的滤波器数量为 8。

(3) 增强滤波器提取特征。空域隐写分析为了丰富残差图像类型,利于隐写特征的提取,其预处理层由一个固定的滤波器逐渐发展为对 SRM 中 30 个高通滤波器的使用。然而,目前的隐写分析模型基本上只是对 SRM 中 30 个滤波器的简单应用,而 SRM 同一类型中的滤波器,只是将提取的不同方向的残差信息作为不同通道的特征图输入卷积层中,尽管提取的残差特征图看起来很全面,但其忽略了不同方向滤波器间的相关性,导致不同方向的残差图是分散的,这对特征提取任务是不利的。类似地,文献[19]和文献[20]在使用导数滤波器和 Gabor 滤波器时也没有考虑这个问题。为了加强不同方向上滤波器的关联,使同一类型滤波器的残差特征更具表现力,提取更加全面的残差特征图,本文对导数滤波器和 Gabor 滤波器的同一类型滤波器的输出特征进行了融合增强。

本文采用平均值融合方法,以导数滤波器为例,通过取权重矩阵在每个方向上的绝对值的平均值来行形成新的权重矩阵。平均值加权可以表示每一方向上共同且相对平均的特征表达。这里考虑到,若对特征图进行相加融合,可能会出现特征权重值一正一负的情况,权重相互抵消,消除了一些重要的隐写特征,干扰了隐写特征的提取;此外,绝对值最大值融合方法只针对滤波后权重矩阵中最显著的特征,造成了一些特征的丢失。因此,本文采用加权平均的特征融合方法来增强滤波器的提取特征。实验结果如表 2 所示,在表中,“Original”“Mean”分别表示原始的 24 个高通滤波器和使用了平均值的融合方法。可以看出,与滤波器的简单使用相比,加入融合增强后的特征图,能帮助模型获取更丰富、更全面的残差信息,提高了模型的检测精度。

表 2 有平均值融合特征和无平均值融合特征的网络在 0.2 bpp 的 WOW 和 S-UNIWARD 两种算法的误检率比较(训练和测试均在 BOSS 数据集上进行)

Table 2 Comparison of false detection rates of 0.2 bpp WOW and S-UNIWARD algorithms for networks with and without mean fusion features. (Training and testing were carried out on the BOSS dataset)

算法	Original	Original + Mean
WOW	0.175 3	0.170 0
S-UNIWARD	0.214 2	0.206 5

2.3 注意力机制

基于卷积神经网络的图像隐写分析主要依靠卷积运算方式,通过融合感受野内的通道信息与空间信息来提取隐写特征。在卷积过程中,所有通道和空间信息都被同等对待,这将导致网络缺乏对关键特征进行判别学习的能力,因此,为了防止网络学习太多无用的特征,影响检测性能,应提高网络对重要特征的关注度。

注意力机制已被证明在提高卷积神经网络的性能方面具有巨大潜力。SENet^[21]是一种挤压和激励的通道注意力,它通过捕捉通道间的相关性,从而为通道分配不同的权重,帮助网络关注那些更重要的特征。为了进一步提升网络性能,有学者提出了通道注意力与空间注意力相结合的注意力模块 CBAM^[22],与 SENet 相比,其在提升网络性能方面更加优异。与上述以更高的模型复杂度为代价来提升模型性能相比,ECA-Net^[23]提出了一种高效通道注意力模块 ECA,该模块实现了以极其轻量级的方式捕捉跨通道交互,同时避免了降维给通道注意力带来的副作用。基于上述考虑,为了在提

升网络性能的同时不增加网络的复杂度,本文采用 ECA 模块替换原始 CBAM 中的通道注意力,将优化后的 CBAM 添加到残差块中,从而帮助网络学习重要特征,抑制不必要特征。

(1)通道注意力。ECA 是一种以低模型复杂度进行有效学习的通道注意力,它通过大小为 k 的快速 1 维卷积有效实现每一通道及其 k 个邻居的局部跨信道交互,并且避免降维对通道注意产生的影响。为了避免手动调整 k 值,该注意力提出了一种自适应确定 k 值的方法,即交互的覆盖范围(内核大小 k)与信道维度成比例。ECA 通道注意力采用了全局平均池化来聚合每一通道信息,这里考虑到最大池化是收集不同特征对象的另一条线索,可以帮助推断更加精细的通道注意力,因此,本文同时使用平均池化和最大池化来聚合特征图的空间信息,以提高网络的表示能力。改善的 ECA 模块(Improved ECA, IECA)结构如图 2 所示。

首先通过平均池化和最大池化操作来聚合特征图的空间信息,生成两个不同的上下文描述符 F_{avg}^c 和 F_{max}^c ,它们分别表示平均池化特征和最大池化特征。然后将这两种特征输入共享网络 1 维卷积中,产生所需要的通道注意力权重 $M_c(F) \in \mathbb{R}^{C \times 1 \times 1}$ 。其中,通过 1 维卷积实现了不降维的局部跨信道策略。其计算过程如下式所示:

$$M_c(F) = \delta(C1D_k(\text{AvgPool}(F))) + \delta(C1D_k(\text{MaxPool}(F))) = \delta(C1D_k(F_{avg}^c)) + \delta(C1D_k(F_{max}^c)) \quad (7)$$

式中: $F \in \mathbb{R}^{C \times H \times W}$ 为输入特征图; δ 为 Sigmoid 激活函数; $C1D$ 为 1 维卷积,参数 k 表示交互的覆盖范围(即 1 维卷积的核大小)。

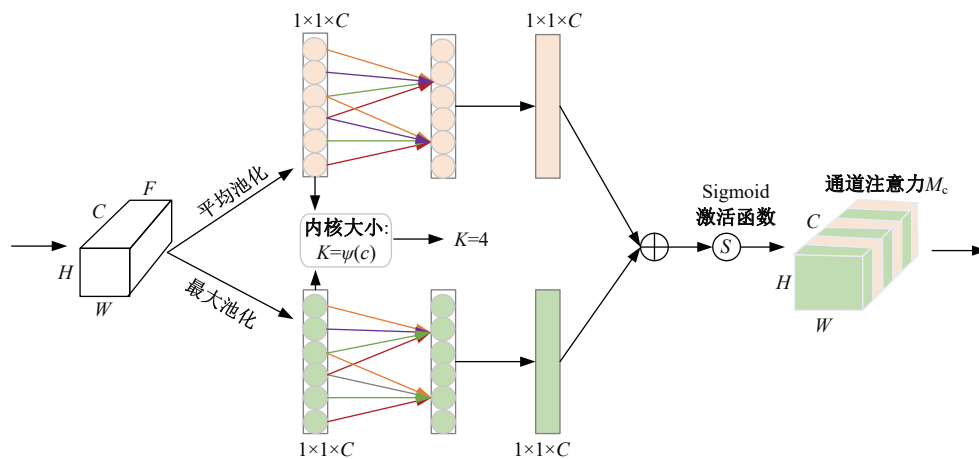


图 2 IECA 模块结构

Fig. 2 IECA module structure

由于ECA模块旨在适当地捕获局部跨通道交互,因此,需要确定交互的覆盖范围,而交互的覆盖范围与通道维度 C 一般是成正比的。换句话说,即 k 与 C 可能存在一个映射 ϕ ,信道维度 C 通常被设置为2的幂,因此,通道维度数量 C 和交互的覆盖范围 k 之间的非线性映射关系可以用下式进行表示:

$$C = \phi(k) = 2^{(\gamma^k - b)} \quad (8)$$

在信道维度 C 已知的情况下,可以通过下式自适应地确定1维卷积的核大小:

$$k = \phi(C) = \left\lceil \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rceil_{\text{odd}} \quad (9)$$

式中: $\lceil \cdot \rceil_{\text{odd}}$ 为 $\cdot$ 的最近奇数。在本文中,所有实验中的 γ 和 b 值分别设置为2和1。

(2)空间注意力。空间注意力的结构如图3所示。其主要功能是通过压缩特征图的通道信息,将整个特征图映射到空间维度,为每个空间像素赋予不同的权重。它更侧重于关注特征图的“哪里”是最具信息量的部分,这与通道注意力形成互补。

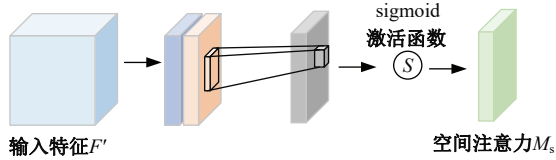


图3 空间注意力的结构图

Fig. 3 Structure of spatial attention

假设一个卷积块的输出为 $F' \in \mathbb{R}^{C \times H \times W}$,其中 H 、 W 和 C 是指特征图高度、宽度和通道数。首先对输入特征图进行平均池化和最大池化操作,获得平均池化特征 $F_{\text{avg}}^s \in \mathbb{R}^{1 \times H \times W}$ 和最大池化特征 $F_{\text{max}}^s \in \mathbb{R}^{1 \times H \times W}$,然后将它们沿通道维度上进行连接,并发送到卷积层生成空间权重图 $M_s(F) \in \mathbb{R}^{H \times W}$ 。简而言之,空间注意力计算过程如式(10)所示:

$$M_s(F) = \delta(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \delta(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (10)$$

式中: δ 为Sigmoid激活函数; $f^{7 \times 7}$ 表示卷积层采用核大小为 7×7 的卷积运算。

(3)空间与通道注意力模块。将IECA通道注意力和空间注意力两个子模块沿着先后顺序串联相结合,形成优化后的注意力模块,将其嵌入残差块中,有效地帮助网络关注那些重要的特征信息,在提高网络性能的同时一定程度地降低模型

复杂度。将优化后的CBAM模块记作CBAM+,其结构如图4所示。

给定一个特征图 $F \in \mathbb{R}^{C \times H \times W}$ 作为输入,特征图 F 经过通道注意模块得到一个1D通道权重图 $M_c \in \mathbb{R}^{C \times 1 \times 1}$,特征图 F 与通道权重 M_c 相乘得到中间特征图 F' ,然后特征图 F' 经过空间注意力模块得到一个2D空间权重图 $M_s \in \mathbb{R}^{1 \times H \times W}$,如图4所示,整个注意力计算过程如式(11)所示:

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (11)$$

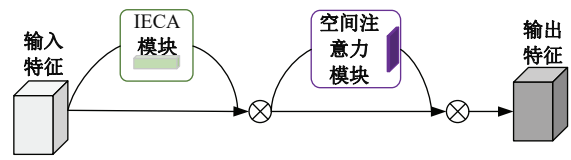


图4 CBAM+结构图

Fig. 4 CBAM+ structure diagram

(4)CBAM+模块的有效性验证。为了验证加入优化后的CBAM+模块对网络性能的提升效果,本文在嵌入率为0.2 bpp和0.4 bpp的S-UNIWARD和WOW隐写算法下,进行了未加注意力模块网络误检率实验和网络分别与CBAM模块及CBAM+模块相结合的对比实验,其实验结果如表3所示。

表3 结合不同注意力模块的网络性能比较(训练和测试均在BOSS数据集上进行)

Table 3 Combined with network performance comparisons of different attention modules (Training and testing were conducted on the BOSS dataset)

模型类型	S-UNIWARD		WOW	
	0.2 bpp	0.4 bpp	0.2 bpp	0.4 bpp
未加注意力机制的模型	0.256 5	0.137 2	0.199 5	0.114 0
添加CBAM的模型	0.207 8	0.120 8	0.170 8	0.099 7
添加CBAM+的模型	0.206 5	0.111 2	0.170 0	0.104 5

如表3所示,加入注意力模块的两种网络其误检测率均低于原始网络,表明加入注意力机制可以帮助网络聚焦图像的隐写区域,增强隐写特征的提取能力,提高网络的性能。此外,相比原始的CBAM注意力模块,优化后的CBAM中的IECA通道注意力可以通过快速1维卷积以极其轻量级的方式,实现每一信道及其 k 个相邻信道的跨信道交互,并有效避免了降维给通道注意力预测带来的副作用。因此,优化后的CBAM注意力模块整体上给网络性能带来了更好的增益。

3 实验结果与分析

本节主要介绍模型的实验设置,包括所使用的数据集、实验平台和模型学习的参数,并以图表的形式展示了本文模型的效果。

3.1 实验设置

实验采用常用的隐写分析图像库 BOSSbase v1.01^[24] 和 BOWS2^[25] 对模型进行性能评估。BOSSbase v1.01 由 10 000 张大小为 512×512 像素的灰度自然图像组成,包含景点、生活、建筑等多类别的图像,是一个隐写分析比赛的专用数据集,图像来源于 7 种不同类型的数码相机,且没有经过任何压缩方式处理。BOWS2 由 10 000 张分辨率为 512×512 像素的灰度图像组成,从未压缩,且具有与 BOSSbase v1.01 数据集相似的内容和特征分布。隐写图像数据集是以原始图像集为嵌入对象,由于计算资源有限,首先将所有图像统一裁剪为 256×256 像素大小,然后采用空域中 3 种常见的自适应隐写算法 S-UNIWARD、WOW、HUGO 在 0.1、0.2、0.3、0.4 四种有效载荷下对这些图像进行信息嵌入生成 stego 图像。将隐写后的每一组图像随机分 7 000 对图像作为训练集,1 000 对图像作为验证集,2 000 对图像作为测试集。

本文实验的软硬件环境如下:操作系统为 Windows 10,32 GB 内存,GPU 型号为 NVIDIA Tesla P40,显存为 24 GB,深度学习框架为 Pytorch(2.1.0)。由于 GPU 内存的限制,将训练批次大小设置为 40(20 个载体/隐写图像对),训练集在每个时期之后打乱增加随机性。采用 Adamax 优化器、交叉熵损失函数,使用 He 初始化器来初始化卷积核权重,所有卷积核的偏置项被初始化为 0.2,对于完全连接层,使用标准差为 0.001 的高斯分布来初始化权重。初始学习率设置为 0.001,并在学习率中加入了预热(warmup)策略^[26],在前 5 个 epochs 进行预热,因此,前 5 个

epochs 的学习率分别为 0.000 2、0.000 4、0.000 6、0.000 8、0.001,此外,在 120 个 epochs 后学习率衰减为 0.000 1,实验总共 250 个 epochs。

隐写分析中,常用的性能评价指标是误检测率,包括虚警率和漏检率。虚警率是指载体图像被判为隐写图像的比率,漏检率是指隐写图像被判为载体图像的比率。假设 TP 表示隐写图像被正确分类的数量,TN 表示载体图像被正确分类的数量,FN 表示隐写图像被误分类为载体图像的数量,FP 表示载体图像被误分类为隐写图像的数量,则虚警率、漏检率及误检率的计算如式(12)(13)(14)所示:

$$P_{FA} = \frac{FP}{FP + TN} \quad (12)$$

$$P_{MD} = \frac{FN}{FN + TP} \quad (13)$$

$$P_E = \min_{P_{FA}} \frac{1}{2} (P_{FA} + P_{MD}) \quad (14)$$

式中: P_{FA} 为虚警率; P_{MD} 为漏检率; P_E 为误检率。

3.2 实验结果

为了进一步验证本文提出的网络结构的性能,选择 Ye-Net、SRNet^[27]、ZhuNet 3 种典型的网络模型与本文方法进行比较,所有的模型都在相同的数据集上进行测试。为了保证实验结果的可靠性,每组实验均重复 3 次,取其均值作为实验结果。实验结果表明:本文方法在精度上有显著提高。

如表 4 和图 5 所示,本文网络相比其他网络具有更好的检测性能。对于不同有效载荷的 S-UNIWARD 算法,本文方法的误检率比 Ye-Net、SRNet 和 ZhuNet 低 0.041 4~0.153 1。对于 WOW 算法所提出的网络与其他网络相比性能也提升了 0.017 2~0.129 9。对于 HUGO 隐写算法的网络的检测精度也提升了 0.024 4~0.134 6。这表明:本文提出的预处理增强和注意力机制的添加可以帮助网络提取更多有用的特征,提高网络的检测准确性。

表 4 不同算法下的隐写分析误检率比较(所有网络都在 BOSS 数据集上进行训练和测试)

Table 4 Comparison of steganographic false detection rates under different algorithms
(All networks were trained and tested on the BOSS dataset)

网络模型	S-UNIWARD				WOW				HUGO			
	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp
Ye-Net	0.446 9	0.358 7	0.286 1	0.237 4	0.393 7	0.291 6	0.245 9	0.213 6	0.427 6	0.347 9	0.284 9	0.256 2
SRNet	0.397 5	0.298 2	0.215 3	0.183 2	0.326 1	0.247 9	0.172 4	0.145 9	0.389 5	0.287 2	0.224 8	0.196 7
ZhuNet	0.366 2	0.258 0	0.185 7	0.152 6	0.315 3	0.221 8	0.156 3	0.121 7	0.377 0	0.264 1	0.206 0	0.166 4
本文算法	0.293 8	0.206 5	0.141 2	0.111 2	0.263 8	0.170 0	0.125 3	0.104 5	0.293 0	0.205 8	0.164 7	0.142 0

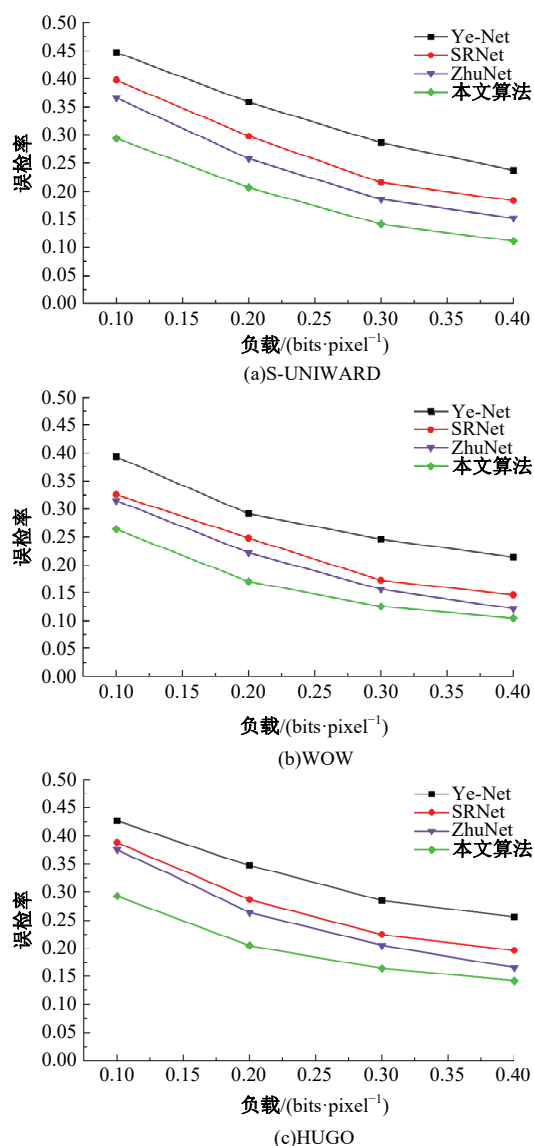


图5 使用Ye-Net、SRNet、ZhuNet和本文算法在S-UNIWARD、WOW和HUGO上,对不同有效载荷隐写分析误检率的比较

Fig.5 Comparison of steganographic false detection rates of different payloads on S-UNIWARD, WOW and HUGO using Ye-Net, SRNet, ZhuNet and the proposed algorithm

在深度学习中,扩大训练数据集的规模,通常训练的模型效果更好。为了研究样本数量的增加对模型性能的影响,本文在BOSS数据集的基础上,加入BOWS2数据集。现在训练数据集包含来自BOSS和BOWS2数据库的14 000对载体/隐写图像,验证集包含2 000对图像,测试集包含4 000对图像。

表5为当数据集规模扩大时,与仅使用BOSS数据集相比,所有网络的检测性能都得到了提升,表明使用规模更大的数据集可以帮助模

表5 BOWS2数据集增强下,不同算法的隐写分析误检率比较

Table 5 Comparison of steganographic error detection rates of different algorithms under BOWS2 data set enhancement

算法	BOSS		BOSS + BOWS2	
	0.2 bpp	0.4 bpp	0.2 bpp	0.4 bpp
Ye-Net	0.291 6	0.213 6	0.256 9	0.142 9
SRNet	0.247 9	0.145 9	0.192 2	0.093 5
ZhuNet	0.221 8	0.121 7	0.173 5	0.087 4
本文算法	0.170 0	0.104 5	0.147 0	0.068 3

型获得更好的性能。

4 结束语

自适应隐写术以不同的概率将秘密信息隐藏在图像的多个区域中,为了帮助网络有效提取隐写特征,本文提出了一种基于预处理层增强和注意力机制的空域图像隐写分析模型。通过引入双重滤波器并对滤波器提取进行融合增强,丰富残差图像类别,增强预处理层的特征提取能力;同时,将优化后的CBAM模块添加到特征提取和特征分析阶段,引导网络关注图像中隐写信号丰富的区域,减小冗余特征的影响。对预处理层和注意力模块的消融实验证明了它们在隐写分析中的有效性。此外,实验结果表明:与其他网络相比,本文方法在检测精度和泛化能力方面都有显著提高。目前,隐写分析算法的复杂度仍然较高,对内存需求较大,未来,将在保证网络性能的同时,进一步优化网络结构,探索适合隐写分析的轻量化网络模型。

参考文献:

- [1] Xiang S J, Luo X R. Reversible data hiding in encrypted image based on homomorphic public key cryptosystem[J]. Journal of Software, 2016, 27(6): 1592-1601.
- [2] 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.
Chen Jun-fu, Fu Zhang-jie, Zhang Wei-ming, et al. A review of image steganography based on deep learning [J]. Journal of Software, 2021, 32(2): 551-578.
- [3] Fridrich J, Kodovsky J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.
- [4] Holub V, Fridrich J. Low-complexity features for

- JPEG steganalysis using undecimated DCT[J]. IEEE Transactions on Information Forensics and Security, 2014, 10(2): 219–228.
- [5] Holub V, Fridrich J. Phase-aware projection model for steganalysis of JPEG images[C]//Media Watermarking, Security, and Forensics, Philadelphia, USA, 2015:75–84.
- [6] Khan A, Sohail A, Zahoor U, et al. A survey of the recent architectures of deep convolutional neural networks[J]. Artificial Intelligence Review, 2020, 53: 5455–5516.
- [7] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//Proc IEEE Conf Comput Vis Pattern Recognit (CVPR), Honolulu, USA, 2017: 1800–1807.
- [8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proc IEEE Conf Comput Vis Pattern Recognit (CVPR), Las Vegas, USA, 2016: 770–778.
- [9] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//AAAI Conf Artif Intell, San Francisco, USA, 2017: 4278–4284.
- [10] Qian Y, Dong J, Wang W, et al. Learning and transferring representations for image steganalysis using convolutional neural network[C]//IEEE International Conference on Image Processing (ICIP), Phoenix, USA, 2016: 2752–2756.
- [11] Jian Y, Ni J, Yang Y. Deep learning hierarchical representations for image steganalysis[J]. IEEE Trans. Inf Forensic Secur, 2017, 12(11): 2545–2557.
- [12] Zhang X B, Zhang X P, Feng G. Image steganalysis network based on dual-attention mechanism[J]. IEEE Signal Processing Letters, 2023, 30: 1287–1291.
- [13] Qian Y, Dong J, Wang W, et al. Deep learning for steganalysis via convolutional neural networks[C]//Media Watermarking, Security, and Forensics, San Francisco, USA, 2015: 171–180.
- [14] Xu G, Wu H Z, Shi Y Q. Structural design of convolutional neural networks for steganalysis[J]. IEEE Signal Processing Letters, 2016, 23(5): 708–712.
- [15] Ye J, Ni J, Yi Y. Deep learning hierarchical representations for image steganalysis[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(11): 2545–2557.
- [16] Zhang R, Zhu F, Liu J, 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.
- [17] Han X, Zhang T. Spatial steganalysis based on non-local block and multi-channel convolutional networks[J]. IEEE Access, 2022, 10: 87241–87253.
- [18] Li B, Li Z, Zhou S, et al. New steganalytic features for spatial image steganography based on derivative filters and threshold LBP operator[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1242–1257.
- [19] 史晓裕, 李斌, 谭舜泉. 深度学习空域隐写分析的预处理层[J]. 应用科学学报, 2018, 36(2): 309–320.
- Shi Xiao-yu, Li Bin, Tan Shun-quan. Preprocessing layer for deep learning spatial steganography[J]. Chinese Journal of Applied Sciences, 2018, 36(2): 309–320.
- [20] Li B, Wei W, Ferreira A, et al. ReST-Net: diverse activation modules and parallel subnets-based CNN for spatial image steganalysis[J]. IEEE Signal Processing Letters, 2018, 25(5): 650–654.
- [21] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132–7141.
- [22] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 3–19.
- [23] Wang Q, Wu B, Zhu P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 11531–11539.
- [24] Bas P, Filler T, Pevný T. Break our steganographic system: the ins and outs of organizing BOSS[C]//International Workshop on Information Hiding, Prague, Czech Republic, 2011: 59–70.
- [25] Bas P, Furon T. BOWS-2. [DB/OL]. [2024-04-12]. <http://bows2.ec-lille.fr>.
- [26] He T, Zhang Z, Zhang H, et al. Bag of tricks for image classification with convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 558–567.
- [27] 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.