

中图法分类号: TP309.7 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Luo Weiwei, Liu Juntong, Gan Boyu. Image steganalysis network integrating equal-resolution frequency bands and contrastive clustering[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.260327. (罗维薇, 刘俊彤, 甘博宇. 融合等分辨率频带与对比聚类的隐写分析网络[J/OL]. 中国图象图形学报, XXXX:1-16. DOI: 10.11834/jig.260327.) [DOI: 10.11834/jig.260327]

融合等分辨率频带与对比聚类的隐写分析网络

罗维薇*, 刘俊彤, 甘博宇

兰州交通大学 电子与信息工程学院, 甘肃省兰州市 730070

摘要: 目的 空域自适应隐写算法通常根据图像内容复杂度选择嵌入位置, 使隐写修改呈现强内容自适应性和稀疏分布特征。由于嵌入扰动幅度较小, 隐写残差容易被图像纹理、边缘和背景噪声掩盖, 现有隐写分析模型在弱隐写信号检测中仍存在频带信息利用不足、局部噪声与全局上下文联合建模不充分、特征空间判别边界不清晰等问题。针对上述问题, 本文提出一种融合等分辨率频带与对比聚类的图像隐写分析网络 CRF-CCNet (Image Steganalysis Network Integrating Equal-resolution Frequency Bands and Contrastive Clustering), 以增强模型对微弱隐写残差的表征和判别能力。方法 在预处理阶段, 本文构建了等分辨率拉普拉斯频带分解模块, 在不进行下采样的条件下提取高频、中频和低频残差信息, 并在各频带上引入 SRM 高通滤波器组, 以增强不同频率成分中的隐写响应; 随后通过自适应注意力融合机制动态调整各频带残差的贡献。在特征提取阶段, 本文设计多路径特征增强模块, 由轻量级 Swin Transformer 路径、深度可分离卷积路径和中央差分卷积路径并行组成, 分别用于建模全局上下文关系、提取局部纹理特征以及增强邻域差值和边缘扰动响应, 并通过门控机制实现多路径特征融合。在分类优化阶段, 引入多尺度对比聚类损失, 通过动态类别原型约束载体图像与隐写图像在特征空间中的分布, 压缩类内距离并扩大类间间隔, 同时利用跨尺度一致性约束保持不同层级隐写特征之间的结构一致性。结果 在 BOSSBase 1.01 和 BOWS2 数据集上, 采用 WOW、HILL 和 S-UNIWARD 三种典型空域自适应隐写算法进行实验。结果表明, 在 BOSS+BOWS2 数据集、0.4 bpp 嵌入率条件下, CRF-CCNet 对 WOW、HILL 和 S-UNIWARD 的检测准确率分别达到 94.81%、89.55% 和 92.47%, 较 HSMNet 分别提高 0.83、1.55 和 1.15%。与轻量级隐写分析网络 LWENet 相比, 本文方法在 WOW0.4bpp 条件下将检测准确率由 90.59% 提高至 94.81%, 提高 4.22%; 同时参数量由 0.82 M 降至 0.67M, 减少约 18.3%。在计算复杂度方面, CRF-CCNet 的 FLOPs 为 20.25 G, 与 LWENet 的 19.45G 处于同一量级, 推理速度达到 82.4FPS。结论 CRF-CCNet 通过等分辨率多频带残差建模、多路径特征增强和多尺度对比聚类优化, 提高了模型对弱隐写扰动的感知能力和特征判别能力。在保持较低参数量和较高推理效率的同时, 所提方法在多种空域自适应隐写算法上取得了较好的检测准确率, 可为轻量化图像隐写分析模型设计提供参考。

关键词: 图像隐写分析; 空域自适应隐写; 等分辨率频带分解; 多路径特征增强; 对比聚类

Image steganalysis network integrating equal-resolution frequency bands and contrastive clustering

Luo Weiwei*, Liu Juntong, Gan Boyu

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

收稿日期: 2026-06-08; 修回日期: 2026-08-23

* 通信作者: 罗维薇, 通信作者, 女, 硕士, 副教授, 主要研究方向为图像处理, 模式识别, 人工智能。E-mail: luoweimei@lztu.edu.cn

基金项目: 国家自然科学基金项目(批准号: 62362047); 甘肃省自然科学基金(批准号: 26JRRA567);

Supported by: Project supported by the National Natural Science Foundation of China (Grant No. 62362047); Project supported by the Natural Science Foundation of Gansu Province, China (Grant No. 26JRRA567)

Abstract: Objective Spatial adaptive steganographic algorithms usually embed secret messages in image regions with complex textures or edges to reduce visual distortion and statistical detectability. Therefore, the resulting embedding modifications are weak, sparse, and highly content-dependent. These subtle traces are easily masked by natural image textures, local intensity changes, and background noise, which makes reliable image steganalysis difficult, especially under low or moderate embedding payloads. Although deep-learning-based steganalysis methods have achieved better performance than handcrafted feature methods, existing models still have several limitations. Most of them rely mainly on single residual preprocessing, which may not fully exploit frequency-band information related to steganographic modifications. In addition, convolutional structures are effective for local texture extraction but have limited ability to model long-range contextual relationships. Moreover, cover and stego features are often highly overlapped in the feature space, making it difficult for cross-entropy loss alone to form a clear decision boundary. To address these problems, this paper proposes an image steganalysis network integrating equal-resolution frequency bands and contrastive clustering, termed CRF-CCNet. **Methods** The proposed CRF-CCNet is designed from three aspects: residual preprocessing, feature extraction, and discriminative optimization. In the preprocessing stage, an equal-resolution Laplacian frequency decomposition module is constructed to extract high-, middle-, and low-frequency residual components without spatial downsampling. This design preserves weak embedding traces while capturing complementary residual responses in different frequency bands. SRM high-pass filter banks are further applied to each band to enhance steganographic residual information. Considering that different frequency bands contribute unequally to detection, an adaptive attention fusion mechanism is introduced to dynamically adjust their weights. In the feature extraction stage, a Multi-Path Feature Enhancement Module is designed with three parallel branches. A lightweight Swin Transformer branch is used to model global contextual dependencies, a depthwise separable convolution branch is adopted to extract local texture features with relatively low computational cost, and a central difference convolution branch is introduced to enhance neighborhood intensity differences and edge-related perturbation responses. The outputs of the three branches are integrated through a gated fusion mechanism, enabling the network to combine global, local, and differential information. In the classification stage, a Multi-Scale Contrastive Clustering Loss is introduced. Dynamic class prototypes are maintained for cover and stego samples at different feature scales, reducing intra-class distances while enlarging inter-class margins. Meanwhile, a cross-scale consistency constraint is imposed on stego features to maintain structural coherence among different feature levels. **Results** Experiments were conducted on the BOSS-Base 1.01 and BOWS2 datasets using three typical spatial adaptive steganographic algorithms, namely WOW, HILL, and S-UNIWARD. The proposed method was compared with representative steganalysis networks, including SRNet, LWENet, FA-CLNet, and HSMNet, under the same data setting. Under the BOSS+BOWS2 training setting with an embedding payload of 0.4 bpp, CRF-CCNet achieves detection accuracies of 94.81%, 89.55%, and 92.47% for WOW, HILL, and S-UNIWARD, respectively, outperforming HSMNet by 0.83, 1.55, and 1.15 percentage points. Compared with the lightweight steganalysis network LWENet, CRF-CCNet improves the accuracy on WOW at 0.4 bpp from 90.59% to 94.81%, with a gain of 4.22 percentage points. Meanwhile, the number of parameters is reduced from 0.82 M to 0.67 M, corresponding to a reduction of approximately 18.3%. Although the FLOPs increase slightly from 19.45 G to 20.25 G, the two models remain at the same computational scale, and CRF-CCNet reaches an inference speed of 82.4 FPS. Ablation experiments further verify the effectiveness of the equal-resolution multi-band preprocessing, the multi-path feature enhancement module, and the multi-scale contrastive clustering loss. **Conclusions** CRF-CCNet improves the representation and discrimination of weak steganographic perturbations by combining equal-resolution multi-band residual modeling, multi-path feature enhancement, and multi-scale contrastive clustering optimization. Experimental results show that the proposed method achieves better detection accuracy for several spatial adaptive steganographic algorithms while maintaining a relatively small parameter size and high inference efficiency. Therefore, CRF-CCNet provides an effective lightweight framework for spatial image steganalysis.

Key words: image steganalysis; spatial adaptive steganography; equal-resolution frequency decomposition; multi-path feature enhancement; contrastive clustering

0 引言

随着社交媒体、云存储和多媒体通信技术的快速发展,数字图像已成为信息传输和信息存储的重要载体。图像隐写术(Image Steganography)通过对载体图像进行细微修改,将秘密信息嵌入其中,并尽量保持图像视觉质量不发生明显变化,从而实现隐蔽通信。该技术在隐私保护、数字版权保护和身份认证等方面具有一定应用价值,但也可能被用于非法隐蔽通信和敏感信息传播,给网络安全监管和数字取证带来潜在风险。因此,判断图像中是否含有秘密信息的图像隐写分析(Image Steganalysis),已成为多媒体安全和数字取证领域的重要研究内容(龙玲慧等,2026)。李文琪等(2025)提出基于扩散模型的生成式视频隐写方法,并通过隐写分析网络验证其安全性,表明新型生成式隐写对隐写分析模型的特征表征与检测适应性提出了更高要求。

早期图像隐写分析方法主要依赖人工设计的统计特征。此类方法通常通过构造残差特征、统计像素邻域关系或建模图像局部差分分布,捕捉隐写嵌入造成的统计异常,并结合支持向量机(Support Vector Machine, SVM)(Suykens 等, 1999)或Fisher线性判别(Fisher Linear Discriminant, FLD)等分类器完成检测(Kodovský 等人, 2012)。在特征设计方面, Pevný 等人(2010)提出 SPAM(Subtractive Pixel Adjacency Matrix)模型,通过统计像素邻域间差分相关性进行隐写检测。Fridrich 等人(2012)进一步提出 SRM(Spatial Rich Model),利用大量高通滤波器提取多方向、多尺度残差统计特征,显著提高了传统隐写分析方法的检测能力。传统统计特征方法具有一定可解释性,在早期隐写算法检测中取得了较好效果,但其性能依赖人工特征设计。当自适应隐写算法不断优化嵌入代价并降低修改痕迹时,固定统计特征难以充分适应更复杂、更隐蔽的嵌入模式。

随着深度学习的发展,卷积神经网络(Convolutional Neural Network, CNN)逐渐被引入图像隐写分析任务。Xu 等人(2016)提出 Xu-Net,将 CNN 用于隐写分析,实现了从图像残差到判别特征的自动学习。Ye 等人(2017)在网络前端引入 SRM 滤波器组,并采用 Truncation Linear Unit(TLU)抑制图像内容信息,使网络更加关注微弱隐写噪声。随后, Boroumand

等人(2019)提出 SRNet,通过深层残差结构增强特征提取能力,使模型能够学习更复杂的隐写模式。相比传统人工特征方法,深度学习模型能够自动提取隐写相关特征,在多种公开数据集和隐写算法上表现出更强的检测能力。与 WOW、HILL、S-UNIWARD 等传统代价函数驱动的自适应隐写方法不同,基于深度学习的修改式隐写方法通常利用神经网络学习嵌入代价、嵌入概率或对抗修改策略,使嵌入位置和修改方式能够根据训练数据自动调整。此类方法在一定程度上突破了人工设计代价函数的限制,但其训练目标、载荷控制方式和隐写图像生成机制与传统空域自适应隐写算法存在差异。部分深度学习隐写方法还与端到端图像隐藏或生成式隐写存在交叉,其隐写痕迹不一定完全表现为传统意义上的局部残差扰动。马宾等人(2024)通过多重对抗与通道注意力提高图像隐写的抗检测安全性。酃妹伊等人(2024)利用纹理生成实现构造式信息隐藏,进一步降低显式修改痕迹。贾孟霖等人(2024)通过超分辨率行为伪装增强图像隐藏的隐蔽性。陈振宇等人(2025)研究密文图像可逆信息隐藏,进一步拓展了信息隐藏的应用场景。因此,针对该类隐写方法的检测通常需要重新构建统一的数据生成流程和评价协议。本文主要关注具有明确载体—含密图像对应关系的空域修改式隐写检测,并以 WOW、HILL 和 S-UNIWARD 作为基准隐写算法开展实验;基于深度学习的修改式隐写检测将在后续工作中进一步研究。

近年来,研究者开始进一步关注轻量化结构、多尺度特征建模以及特征优化等问题。Chen 等人(2021)提出结合轻量化结构与多尺度二阶统计特征的隐写分析方法,在控制模型复杂度的同时增强对复杂隐写噪声的表征能力。部分研究借鉴 DenseNet 结构(Huang 等, 2017)和 EfficientNet 结构(Tan 等, 2019),通过增强特征传递和结构复用提高网络表达能力。多尺度卷积、空洞卷积和特征金字塔等结构也逐渐被用于隐写分析,以增强模型对不同尺度嵌入扰动的感知能力。LWENet(Weng 等, 2022)通过轻量级网络设计降低参数量和计算复杂度,在轻量化隐写分析方面取得了较好效果。Fu 等人(2025)提出基于双路径增强与分形下采样的隐写分析网络,通过双路径特征增强和分形下采样结构提升特征表达能力。Ma 等人(2025)提出 EIS-OBEA 方法,

在特征提取过程中动态筛选关键隐写特征子模型,以降低特征冗余和计算复杂度。

除卷积结构外,注意力机制也被用于增强隐写分析模型对关键区域和关键通道的关注能力。Hu 等人(2018)提出 SE(Squeeze-and-Excitation)模块,通过通道重标定增强重要特征响应。Peng 等人(2019)提出双重注意力网络,通过联合建模空间依赖关系与通道依赖关系提高特征表达能力。在隐写分析任务中,注意力机制能够帮助模型抑制图像内容带来的干扰,并突出与隐写修改相关的弱响应区域。罗维薇等人(2025)提出基于预处理层增强和注意力机制的空域图像隐写分析方法,通过增强预处理阶段的隐写残差响应,并结合注意力机制对关键特征区域进行建模。HSMNet(Yang 等,2026)提出一种基于混合空洞卷积与自注意力多通道网络的灰度图像隐写分析方法,通过扩大感受野并融合不同尺度特征,提高了模型对复杂隐写的检测性能。

同时,Transformer 结构因其全局依赖建模能力,也逐渐被引入图像隐写分析。与卷积结构主要关注局部邻域不同,Transformer 能够建立远距离像素或区域之间的依赖关系,有助于描述图像中的全局统计变化。Luo 等人(2022)将卷积视觉 Transformer 用于图像隐写分析,通过结合卷积结构与 Transformer 全局建模能力,提高了模型对局部纹理特征和全局空间关系的联合表征能力。He 等人(2025)提出基于轻量级 DenseLite 和自适应注意力机制的隐写分析网络,通过密集连接和门控融合增强对隐写敏感区域的感知能力。Yan 等人(2025)提出 FA-CLNet,通过残差聚焦增强模块与自适应原型对比学习模块联合增强隐写残差特征表达和类间判别能力。He 等人(2025)提出 ESNNet,通过双分支预处理实现空间域与 JPEG 域隐写信号的统一建模,在不同隐写域上取得了较好的检测性能。

尽管上述方法推动了深度学习隐写分析的发展,但针对空域自适应隐写的弱修改特征,现有模型仍存在一些值得进一步研究的问题。首先,许多方法在预处理阶段仍主要依赖单一高通残差。高通滤波能够增强局部嵌入扰动,但空域自适应隐写并不是简单地在高频区域产生固定模式的噪声,而是根据图像内容复杂度调整嵌入位置和修改代价。隐写扰动在不同频率成分中的表现并不完全一致,部分弱修改可能体现为纹理过渡、局部结构或低幅度残

差变化。若仅使用单一残差表示,模型容易忽略不同频带之间的互补信息。其次,现有多路径或多分支结构虽然已经被用于隐写分析,但多数工作更强调网络宽度或特征并联,对不同路径之间“各自提取什么隐写信息、如何避免冗余、如何按图像内容动态融合”考虑不足。对于弱隐写信号而言,单纯堆叠多个分支并不一定带来稳定提升,关键在于使不同路径分别承担明确的表征功能,并通过自适应机制选择有效响应。最后,载体图像与隐写图像之间的差异极其微弱,两类样本在特征空间中往往高度重叠。仅使用交叉熵损失进行分类训练,虽然能够完成基本判别,但对类内紧凑性和类间可分性的约束不够直接,容易导致判别边界不够清晰。

基于上述分析,本文提出一种融合等分辨率频带与对比聚类的图像隐写分析网络 CRF-CCNet(Image Steganalysis Network Integrating Equal-resolution Frequency Bands and Contrastive Clustering)。本文的出发点并不是简单增加滤波器或叠加多条特征路径,而是围绕弱隐写残差的形成特点,从频带残差构造、噪声特征协同表达和特征空间判别约束三个层面重新组织网络。在预处理阶段,本文构建等分辨率频带残差建模方式,通过连续平滑与差分操作获得高频、中频和低频残差,并保持各频带特征与原图相同的空间分辨率,避免下采样造成弱残差信息丢失。SRM 高通滤波器作为稳定的残差增强算子,用于对不同频带残差进行方向性响应提取;设计重点在于等分辨率频带残差构造、残差归一化以及不同频带响应的自适应融合,使模型能够根据图像内容动态利用有效频带信息。

在特征提取阶段,本文设计面向隐写弱扰动的异构路径协同建模模块多路径特征增强 MP-FEM 模块(Multi-Path Feature Enhancement Module, MP-FEM)。与一般多路径结构不同,该模块并不简单追求分支数量,而是让不同路径分别对应隐写分析中常见的三类线索:Transformer 路径用于补充卷积结构难以充分建模的长距离上下文关系,深度可分离卷积路径用于提取局部纹理变化,中央差分卷积路径用于增强邻域灰度差异和边缘微扰响应。随后通过自适应门控融合机制对三类特征进行动态选择,使网络能够根据输入图像中隐写痕迹的表现形式调整不同路径的贡献。既保留了多路径特征的互补性,减少了无效特征堆叠带来的冗余。

在分类优化阶段,本文引入多尺度对比聚类损失 MS-CCL(Multi-Scale Contrastive Clustering Loss)。该损失从低、中、高不同层级特征中构建类别原型,使载体图像和隐写图像在多尺度特征空间中形成更加稳定的聚类结构。考虑到隐写修改在不同感受野下应具有一定结构一致性,本文进一步对隐写样本施加跨尺度一致性约束,以减少不同层级特征之间的判别冲突。与仅依赖交叉熵损失相比,该策略能够更加直接地压缩类内距离、扩大类间间隔,从而增强模型对弱隐写样本的区分能力。

本文的主要工作如下:

1)提出一种等分辨率频带残差建模机制,通过无下采样的多频带残差构造、归一化与自适应融合,增强不同频带弱隐写信息的互补表征。需要说明的是,SRM滤波器在该模块中主要作为稳定的方向性残差增强算子使用,本文的创新重点不在于重新设计SRM滤波器,而在于等分辨率频带残差构造、弱残差信号保持以及频带响应的自适应融合。

2)提出一种面向弱隐写扰动的功能分工式异构路径协同建模机制。已有多分支隐写分析方法通常侧重于增加网络宽度或并联特征提取路径,而本文进一步明确不同路径的隐写分析功能分工:Transformer路径用于建模长距离上下文依赖,深度可分离卷积路径用于提取局部纹理扰动,中央差分卷积路径用于增强邻域灰度差异和边缘微扰响应。随后,利用门控融合机制根据输入样本的特征分布动态分配三类路径权重,从而减少无效分支堆叠带来的冗余,提高对微弱隐写痕迹的针对性表征能力。

3)提出一种面向特征重叠问题的多尺度对比聚类约束。针对低嵌入率下两类样本在特征空间中高度重叠的问题,本文在低、中、高不同层级特征上构建动态类别原型,通过原型对比约束压缩类内距离并扩大类间间隔。同时,考虑到隐写扰动在不同感受野尺度下应具有一定一致性,本文仅对隐写样本施加跨尺度一致性约束,以减少不同层级隐写特征之间的判别冲突,进一步增强模型对弱隐写样本的区分能力。

1 模型设计

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。等分辨率频带分解增强不同尺度隐写残差;MP-FEM模块负责跨尺度特征建模;MS-CCL进一步约束不同尺度特征空间分布,三者形成了频带增强、特征建模、判别优化的完整网络。

1.1 网络架构

研究提出的总体网络结构如图1所示,网络接收大小为 256×256 的单通道灰度图像作为输入,并最终输出图像是否包含隐秘信息的二分类判定。网络主要由预处理、特征提取与分类判别三个阶段组成。首先,等分辨率多频带预处理模块通过连续平滑与差分操作构造高频、中频和低频残差,并利用固定SRM高通滤波器组对不同频带残差进行方向性增强;随后,自适应频带融合机制根据不同频带的响应强度动态调整其贡献。其次,特征提取阶段通过MP-FEM模块构建Transformer路径、卷积路径和中央差分卷积路径,分别提取上下文依赖、局部纹理和邻域差分扰动信息,并通过Softmax门控机制进行自适应融合。最后,分类阶段在交叉熵损失基础上引入多尺度对比聚类损失,以增强载体图像与隐写图像在特征空间中的可分性。

1.2 等分辨率拉普拉斯分解预处理模块

等分辨率拉普拉斯分解预处理模块如图2所示。该模块的目的是通过连续的高斯模糊算子提取多频带残差,但始终保持特征图分辨率与输入一致(Zero-loss),从而完整保留了在不同嵌入率下的隐写特征。随后,利用稳定频带融合模块(Stable Band Fusion),根据不同频带的响应强度,自适应地调整高、中、低频信号的贡献比例。

本文构建的低频支路并不是直接将原始低频图像内容输入高通滤波器,而是将相邻平滑尺度之间的差分结果作为低频残差表示。具体而言,输入图像首先经过连续高斯平滑获得不同尺度的平滑表示,然后通过相邻尺度之间的差分构造高频、中频和低频残差。由于三类残差均保持与输入图像相同的空间分辨率,因此能够在不破坏像素对应关系的前提下保留不同尺度下的弱隐写响应。记高频、中频和低频残差分别为 h_1 、 h_2 、 h_3 ,其空间尺寸均为 $H \times W$ 。

在此基础上,固定SRM高通滤波器组被用作方向性残差增强算子,分别作用于不同频带残差,而不是重新设计新的SRM滤波器。低频残差支路的作用是从较平滑结构变化中提取可能与隐写扰动相关

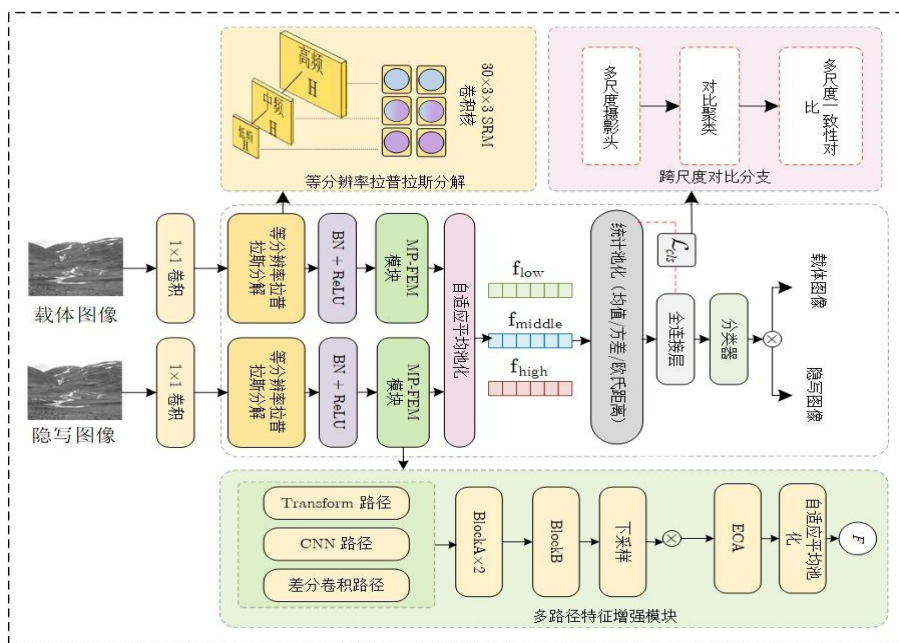


图1 融合等分辨率频带与对比聚类的图像隐写分析网络框架

Fig. 1 Image Steganalysis Network Combining Equal-resolution Frequency Bands and Contrastive Clustering

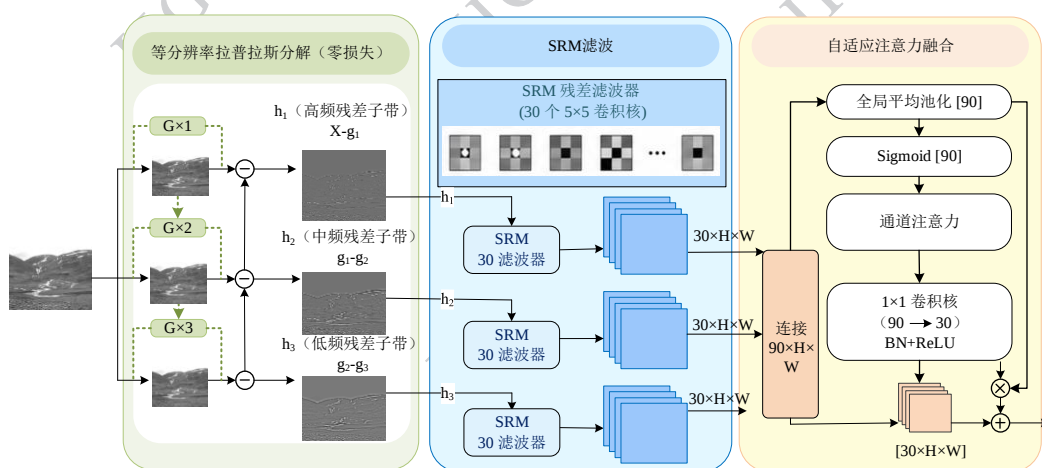
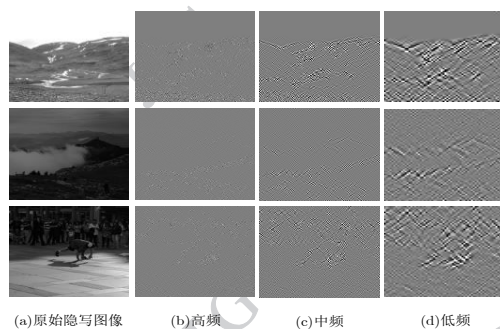


图2 等分辨率拉普拉斯分解预处理模块图

Fig. 2 Equal-resolution Laplacian decomposition preprocessing framework

的弱响应。当低频残差中有效判别信息较少时,自适应门控机制会降低该支路权重;当低频结构中仍包含有助于区分载体图像和隐写图像的微弱扰动时,该支路可为最终分类提供补充信息。图3给出了输入图像经过等分辨率拉普拉斯分解后得到的不同频带残差可视化结果。

然后进行残差自适应归一化。由于不同频带的隐写噪声能量差异极大,为防止弱信号在后续卷积中被湮灭,对提取的残差特征 h_k 进行基于绝对值均值的归一化处理,并引入可学习的频带尺度因子 s_k :



((a)Original-Image; (b)High-band; (c)Mid-band; (d)Low-band)

图3 等分辨率拉普拉斯分解图

Fig. 3 Equal-resolution Laplacian Decomposition Diagram

$$\hat{h}_k = \frac{h_k \cdot s_k}{\frac{1}{H \times W} \sum_{i,j} |h_k(i,j)| + \varepsilon}, \quad k \in \{1,2,3\} \quad (1)$$

式中, h_k 表示第 k 个频带的原始残差特征, $\hat{h}_k$ 表示归一化后的残差特征, s_k 为第 k 个频带的可学习尺度因子, $k \in \{1,2,3\}$ 分别对应高、中、低频带, i, j 表示像素空间位置, H 和 W 分别表示特征图的高度和宽度, ε 为防止分母为零的极小常数。

最后进行不同频带残差上的固定 SRM 方向性增强与自适应软门控融合。将归一化后的残差分别送入固定的 SRM 滤波器组, 提取具有指向性的隐写噪声特征, 为了动态平衡不同频带对最终判别的贡献, 通过全局平均池化 (Global Average Pooling, GAP) 和多层感知机 (Multi-Layer Perceptron, MLP) 以及 σ 激活函数, 聚合多频带的全局上下文, 生成自适应门控权重 G :

$$G = \sigma \left(\text{MLP} \left(\text{GAP} \left(S_1 \parallel S_2 \parallel S_3 \right) \right) \right) \in \mathbb{R}^3 \quad (2)$$

最终, 以高频残差 S_1 为基准特征, 通过门控权重对中频 S_2 和低频 S_3 进行动态加权融合。 $\parallel$ 表示特征拼接操作, $\sigma(\cdot)$ 表示 Sigmoid 激活函数, $\mathbb{R}^3$ 表示门控权重, G 为 3 维向量。

1.3 特征提取模块

特征提取模块如图 4 所示。本阶段的目标是将预处理阶段获得的频带残差进一步转换为具有判别性的深层特征。网络首先利用多路径特征增强模块 (MP-FEM) 对输入残差进行协同建模, 并通过自适应门控机制完成特征融合; 随后, 融合特征进入由残差块和通道注意力模块组成的主干网络, 逐步完成特征降维、层级抽象和多尺度判别特征输出。其中, B, C, H, W 分别表示批量大小、通道数、特征图高度和宽度; w 表示 Transformer 的窗口尺寸。

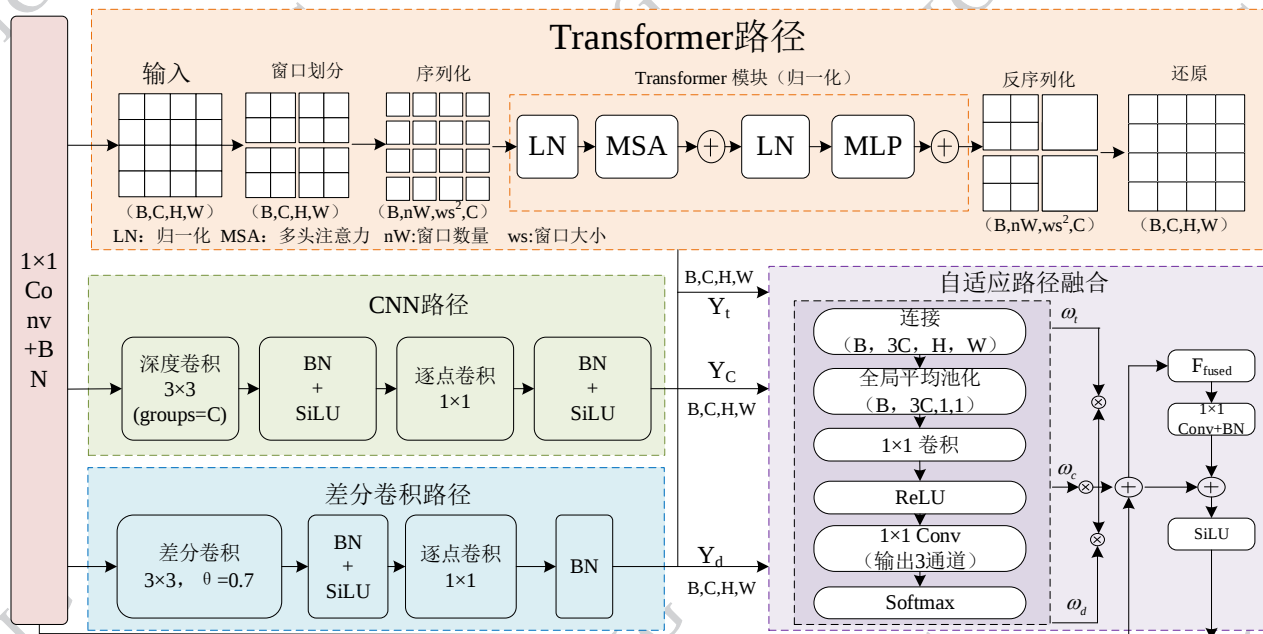


图 4 特征提取阶段流程图

Fig. 4 Flowchart of Feature Extraction Stage

MP-FEM 模块用于解决单一路径结构难以同时刻画上下文依赖、局部纹理和邻域差分扰动的问题。本文采用多路径结构并非简单增加并行分支, 而是针对弱隐写扰动的不同表现形式进行功能分工: Transformer 路径侧重建模较大范围的上下文依赖, CNN 路径侧重提取局部纹理变化, 中央差分卷积路径侧重增强邻域灰度差异和边缘微扰响应。三类特征随后通过门控机制进行自适应融合, 使网络能够

根据不同图像内容和隐写痕迹分布选择更有效的判别响应。

1.3.1 多路径特征增强模块

MP-FEM 模块由三条并行路径协同构成:

1) Transformer 路径: 用于建模较大范围内的上下文依赖关系, 补充卷积结构局部感受野受限的问题。采用轻量级 SwinTransformer 子模块作为核心组件, 通过窗口注意力机制实现较大范围上下文依赖。

核心窗口自注意力计算公式为:

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} + \mathbf{M}\right)\mathbf{V} \quad (3)$$

$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 表示注意力函数,用于计算特征序列的加权表示;式中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别为查询矩阵(Query)、键矩阵(Key)和值矩阵(Value),三者均由输入特征经线性变换得到。 $\mathbf{Q}\mathbf{K}^T$ 为查询与键的点积相似度矩阵,反映不同位置特征之间的相关性;为避免点积值过大导致 Softmax 函数梯度消失,需将相似度矩阵按键向量维度 d_k 的平方根进行缩放。 $\mathbf{M}$ 为掩码矩阵,用于屏蔽序列中的无效位置使模型仅关注有效上下文。最后,将该权重矩阵与值矩阵 $\mathbf{V}$ 相乘,得到加权后的注意力输出特征。

2) CNN 路径:采用深度可分离卷积可以提取像素级局部纹理特征,聚焦隐写噪声对应的细粒度纹理异常。深度可分离卷积在降低了参数量的前提下,有助于捕捉局部像素的纹理突变,适配隐写噪声的细粒度特性,使其为局部特征投影提供核心的纹理信息。CNN 路径采用深度可分离卷积(DepthwiseSeparableConvolution)构建残差结构,平衡特征提取能力与计算复杂度,深度卷积(DepthConv)对每个输入通道独立卷积,捕捉单通道内的局部纹理;点卷积(PointConv)通过 1×1 卷积实现通道间特征融合。

3) 中央差分增强路径:采用中央差分卷积(Central Difference Convolution, CDC)替代传统卷积,通过模拟梯度计算强化噪声对应的像素差值信号,直接对像素差值建模,能够有效放大隐写噪声导致的像素值微小变化,同时抑制平滑背景区域的有效响应,提升噪声特征的信噪比。公式为:

$$\mathbf{Y}_{\text{diff}}(\mathbf{p}_0) = \sum_{\mathbf{p}_n \in \Omega} w_n [\mathbf{X}(\mathbf{p}_0 + \mathbf{p}_n) - \mathbf{X}(\mathbf{p}_0)] \quad (4)$$

$\mathbf{Y}_{\text{diff}}(\mathbf{p}_0)$ 表示像素 $\mathbf{p}_0$ 的加权差值特征, $\mathbf{X}(\mathbf{p}_0)$ 表示像素 $\mathbf{p}_0$ 的原始像素值, $\mathbf{p}_n$ 为以 $\mathbf{p}_0$ 为中心的邻域内的第 n 个像素位置, $\mathbf{X}(\mathbf{p}_0 + \mathbf{p}_n)$ 为邻域像素 $\mathbf{p}_0 + \mathbf{p}_n$ 的像素值。 w_n 为对应邻域像素的权重系数,用于控制不同位置邻域像素对中心像素的贡献程度, Ω 表示以中心位置 $\mathbf{p}_0$ 为中心的邻域位置集合。

1.3.2 自适应权重融合机制

为了动态调整不同路径对最终特征表示的贡献,本文引入基于 Softmax 门控的自适应融合机制。首先,对三条路径输出的特征进行拼接,并利用全局

平均池化获得全局描述向量;随后,通过轻量级门控网络生成三条路径的权重系数,并使其满足: $i \in \{t, c, d\}$, $\sum_{i \in \{t, c, d\}} w_i = 1$;最终融合特征通过加权求和与全局残差连接获得:

$$\mathbf{F} = \text{SiLU}(\text{Proj}(w_t \mathbf{Y}_t \oplus w_c \mathbf{Y}_c \oplus w_d \mathbf{Y}_d) + \mathbf{X}) \quad (5)$$

$\mathbf{F}$ 表示多分支特征融合后的输出特征; $\mathbf{Y}_t, \mathbf{Y}_c, \mathbf{Y}_d$ 分别为不同分支的输入特征, w_t, w_c, w_d 为各分支对应的可学习权重,用于控制不同特征对融合结果的贡献。Proj($\cdot$) 为投影层,用于将融合后的特征映射到统一的特征空间保证特征维度一致。最后通过残差连接与输入特征 $\mathbf{X}$ 相加,得到融合特征。 t, c, d 分别表示 Transformer、CNN 和 CDC 路径, $\oplus$ 表示特征拼接操作, SiLU 表示激活函数。

1.3.3 分层多尺度特征输出

在 MP-FEM 模块之后,特征流经交替堆叠的残差块与通道注意力层,完成特征在不同感受野尺度下的降维与特征提取,最终输出三类判别特征。

1.4 分类模块

多尺度对比聚类损失所在的分类模块如图 5 所示。由于载体图像与隐写图像之间的嵌入扰动十分微弱,两类样本在特征空间中往往存在较大重叠,且同一类别样本在不同尺度特征上的分布也可能较为分散。仅依赖交叉熵损失虽然能够完成基本分类,但其主要关注样本是否被正确划分,对类内特征是否紧凑、类间距离是否充分拉开缺乏直接约束。因此,本文引入基于动态类别原型的多尺度对比聚类约束,通过为载体类和隐写类分别构建稳定的类别中心,引导同类样本向对应原型聚集,并推动异类样本远离对应原型,从而压缩类内距离、扩大类间间隔,提高模型对微弱隐写噪声的判别能力。

1.4.1 多尺度特征提取与投影

在分类阶段前,模型首先从主干网络中提取局部、中尺度、全局三个层级的特征,分别对应隐写噪声的细粒度隐写信息、中等粒度隐写纹理信息与全局隐写信息。为实现多尺度特征在同一空间的对比学习,设计三个独立的特征投影头,将不同维度的特征映射至 128 维公共特征空间并进行 L_2 归一化。

1.4.2 动态类别原型构建

本文在每一尺度上分别维护载体类和隐写类两个原型向量,并采用动量更新机制对其进行更新,利用当前批次(Batch)内同类样本的中心特征对原型

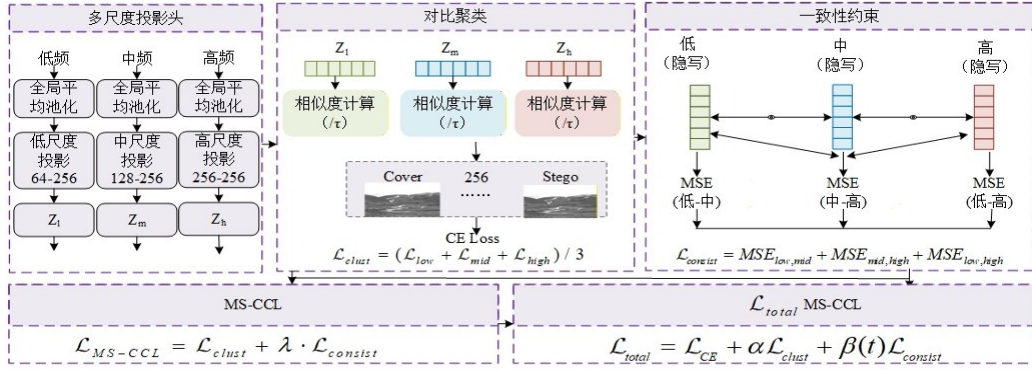


图5 多尺度对比聚类损失模块图

Fig. 5 Multi-scale Contrastive Clustering Loss Module Diagram

进行平滑平移,确保聚类中心的稳定性:

$$\mathbf{P}_c^{(t)} = N \left(m \cdot \mathbf{P}_c^{(t-1)} + (1 - m) \cdot \bar{\mathbf{Z}}_c^{(t)} \right) \quad (6)$$

$\mathcal{P}_c^{(t)}$ 表示更新后的第 c 类中心,它由上一轮的旧中心 $\mathcal{P}_c^{(t-1)}$,和当前批次数据计算出的临时中心 $\bar{\mathbf{Z}}_c^{(t)}$ 加权得到。动量系数 m 控制两者的权重。最后通过归一化操作 N 将更新后的中心调整到统一尺度,保证后续相似度计算公平性。 c 表示类别索引,本文 $c \in \{\text{cover}, \text{stego}\}$; t 表示当前原型更新迭代次数。

1.4.3 多尺度对比聚类损失计算

MS-CCL 损失由尺度内对比损失与跨尺度一致性约束两部分构成,总损失函数定义为:

$$L_{\text{MS-CCL}} = L_{\text{clust}} + \lambda L_{\text{consist}} \quad (7)$$

式中, λ 为跨尺度一致性损失的权重系数,用于平衡尺度内聚类约束与跨尺度一致性约束。

尺度内聚类损失 $\mathcal{L}_{\text{clust}}$: 利用温度系数 $\tau = 0.1$ 缩放特征与原型的相似度,通过交叉熵形式强制样本向同类原型聚集,在计算时 $\mathcal{L}_{\text{clust}}$ 对原型矩阵进行 detach 处理,防止原型在反向传播中被错误梯度破坏:

$$L_{\text{clust}} = -\frac{1}{3} \sum_{s \in \{l, m, h\}} \log \frac{\exp(\mathbf{z}_s^T \mathbf{p}_y / \tau)}{\sum_{c \in C} \exp(\mathbf{z}_s^T \mathbf{p}_c / \tau)} \quad (8)$$

$\mathcal{L}_{\text{clust}}$ 表示最终的聚类损失值; $\mathbf{z}_s$ 表示不同尺度或分支的样本特征, $s \in \{l, m, h\}$ 表示对三种不同层级的特征都计算损失。分子计算样本特征与正确类别中心的相似度,分母计算样本特征与所有类别中心的相似度, y 表示当前样本的真实类别标签, $\mathbf{p}_y$ 表示其对应的正确类别原型, $\mathbf{p}_c$ 表示第 c 类原型, c 为类别索引, C 表示类别集合。

跨尺度一致性约束 $\mathcal{L}_{\text{consist}}$: 基于隐写噪声在多尺

度下印记一致的物理特性,仅对隐写样本施加不同尺度之间的两两均方误差 (Mean Squared Error, MSE) 约束,消除不同感受野下的特征冲突:

$$L_{\text{consist}} = \frac{\sum_{i \in \mathcal{D}_{\text{stego}}} [M(\mathbf{z}_l^i, \mathbf{z}_m^i) + M(\mathbf{z}_m^i, \mathbf{z}_h^i) + M(\mathbf{z}_l^i, \mathbf{z}_h^i)]}{|\mathcal{D}_{\text{stego}}|} \quad (9)$$

式中, $\mathcal{D}_{\text{stego}}$ 表示当前批次中的隐写样本集合, $|\mathcal{D}_{\text{stego}}|$ 表示隐写样本数量, i 为样本索引, $\mathbf{z}_l^i, \mathbf{z}_m^i, \mathbf{z}_h^i$ 分别表示第 i 个隐写样本在低、中、高尺度上的特征表示, M 表示均方误差函数。

1.4.4 联合损失优化

最终优化目标由主分类交叉熵损失 $\mathcal{L}_{\text{CE}}$ 、尺度内对比聚类损失 $\mathcal{L}_{\text{contrast}}$ 和跨尺度一致性约束共同组成。式中, α 为对比损失权重, $\beta(t)$ 为随训练过程动态变化的 MS-CCL 权重, t 表示当前训练轮数:

$$L_{\text{total}} = L_{\text{CE}} + \alpha \cdot L_{\text{contrast}} + \beta(t) \cdot L_{\text{MS-CCL}} \quad (10)$$

2 实验设置

2.1 数据集

在实验中,使用了两个在隐写分析领域常用的数据集: BOSSBase1.01 和 BOWS2。每个数据集包含 1 万张大小为 512×512 的图像,我们统一将图像大小调整为 256×256 。为了研究提出的模型在不同数据集下的性能,我们将数据分为两个训练场景:

1) 将训练集 BOSSBase1.01 数据集按比例分为训练集、验证集和测试集的 6:1:3,即训练集中有 6000 对 cover/stego 对,验证集中有 1000 对图像对,测试集中有 3000 对图像对。

2) 训练集 BOSS+BOWS2: 在训练集 BOSS 的基础上,从 BOWS2 数据集中增加 10000 对 cover/stego

对。即训练集中有 16000 个图像对,验证集中有 1000 个图像对,测试集中有 3000 个图像对。

2.2 实验环境

对于隐写分析,使用 S-UNIWARD (Holub 等, 2014)、HILL (Li 等, 2014) 和 WOW (Holub 等, 2012) 三种常用的隐写算法构建了隐写数据集,通过 Matlab 将秘密信息随机密钥嵌入到图像中。为了验证所提出的 CRF-CCNet 的有效性,将其与两种经典的网络和两种最新的网络进行了比较,包括 SRNet (Boroumand 等, 2019)、LWENet (Weng 等, 2022)、FA-CLNet (Yan Li 等, 2025)、HSMNet (Yang 等, 2025)。为了保证实验的公平性,所有隐写分析网络使用的训练集、验证集和测试集都是相同的。隐写分析网络均采用 Pytorch1.11.0 框架实现,实验均在服务器 3090 上运行。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。

2.3 超参数设置

在训练阶段,将 mini-batch 大小设置为 32 (16 对 cover/stego),并将 cover 和 stego 图像作为成对进行训练。实验采用了与 SRNet 训练相同的数据增强技术 (随机镜像和旋转)。我们使用随机梯度下降 (Stochastic Gradient Descent, SGD) 来优化网络参数,初始学习率为 0.01。该网络总共训练了 500 个 epoch,在第 299 和 399 个 epoch 的学习率除以 10。用来获得最优化的模型。

为降低过拟合风险并增强模型对图像几何变换的鲁棒性,训练阶段采用随机旋转与镜像的数据增强策略。其过程可表示为:

$$I_{aug} = T(I), \quad T \in D_4 \quad (11)$$

I 为输入图像, I_{aug} 为增强后的图像 D_4 为由旋转

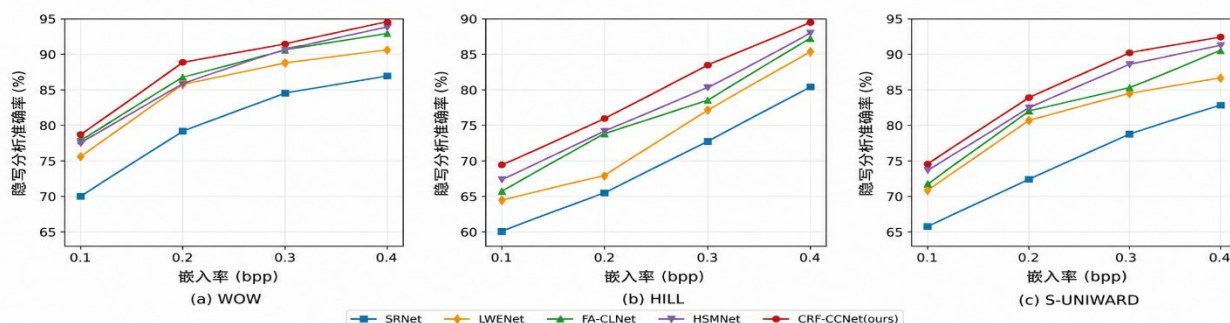
和镜像操作组成的二面体变换群, T 表示从二面体变换群 D_4 中随机选取的几何变换操作。

2.4 与其他方法的比较

表 1 给出了本文方法与 SRNet、LWENet、FA-CLNet 和 HSMNet 在 BOSS+BOWS2 数据集上的检测结果。5 种模型在 BOSS+BOW2 上不同嵌入载荷下不同隐写方法检测准确率 ACC (Accuracy) (%) 和曲线下面积 AUC (Area Under the ROC Curve, AUC) (%) 的比较,如图 6 所示。

从低嵌入载荷结果来看,在 0.1 bpp 和 0.2 bpp 条件下,CRF-CCNet 在 WOW、HILL 和 S-UNIWARD 三种隐写算法上整体保持较高检测准确率,说明所提出的等分辨率频带残差建模和多尺度对比聚类约束能够在隐写扰动较弱时仍提供有效判别信息。从高嵌入载荷结果来看,在 0.3 bpp 和 0.4 bpp 条件下,随着嵌入信息量增加,不同模型的检测准确率整体均有所提升,但 CRF-CCNet 在三种隐写算法下仍保持最优或较优性能。尤其在 0.4 bpp 条件下,CRF-CCNet 对 WOW、HILL 和 S-UNIWARD 的检测准确率分别达到 94.81%、89.55% 和 92.47%,均高于对比模型,说明本文方法不仅能够增强低嵌入载荷下微弱隐写扰动的感知能力,也能够在中高嵌入载荷条件下保持稳定的特征判别优势。综合不同嵌入载荷下的实验结果可知,等分辨率频带残差建模、多路径特征增强和多尺度对比聚类约束能够从残差保持、特征协同和判别优化三个层面共同提升模型对空域自适应隐写痕迹的检测能力。

图 7 给出了五种模型在 WOW 0.4 bpp 条件下的 ROC (Receiver Operating Characteristic, ROC) 曲线。可以看出,CRF-CCNet 的 ROC 曲线整体位于其



((a)WOW; (b)HILL; (c)S-UNIWARD)

图 6 在 BOSS+BOW2 下不同嵌入载荷下不同隐写方法检测准确率 (%) 的比较

Fig. 6 Comparison of Detection Accuracy (%) for Different Steganography Methods under Various Embedding Payloads

表 1 提出方法与以往方法在四种有效载荷和三种隐写方法的比较

Table 1 Comparison of the proposed method with existing methods under four payloads and three steganography algorithms

Algorithm	Arithmetic	0.1		0.2		0.3		0.4	
		ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
WOW	SRNet	69.84	82.83	79.11	85.65	84.42	88.30	86.87	91.58
WOW	LWENet	75.67	83.97	85.78	87.85	88.82	89.97	90.59	95.36
WOW	FA-CLNet	77.69	85.26	86.88	89.89	90.65	91.44	93.05	96.61
WOW	HSMNet	77.35	86.12	86.00	90.37	90.74	92.22	93.98	97.84
WOW	CRF-CCNet(ours)	78.51	86.57	88.91	91.00	91.47	93.05	94.81	98.75
HILL	SRNet	60.20	71.17	65.58	75.16	72.74	79.67	80.39	88.02
HILL	LWENet	64.53	71.39	68.07	76.60	77.08	80.58	85.42	90.43
HILL	FA-CLNet	65.78	72.36	73.84	77.44	78.49	83.14	87.37	92.63
HILL	HSMNet	67.41	72.42	74.25	77.97	80.31	85.85	88.00	94.92
HILL	CRF-CCNet(ours)	69.52	73.64	75.93	79.39	83.42	86.03	89.55	95.59
S-UNIWARD	SRNet	65.75	74.61	72.45	80.18	78.69	88.16	82.91	90.15
S-UNIWARD	LWENet	70.89	76.07	80.74	82.41	84.54	89.21	86.68	91.12
S-UNIWARD	FA-CLNet	71.72	80.27	82.04	86.30	85.26	90.90	90.65	94.84
S-UNIWARD	HSMNet	73.69	81.64	82.45	88.68	88.64	90.31	91.32	95.26
S-UNIWARD	CRF-CCNet(ours)	74.54	82.31	83.89	89.52	90.38	91.85	92.47	97.38

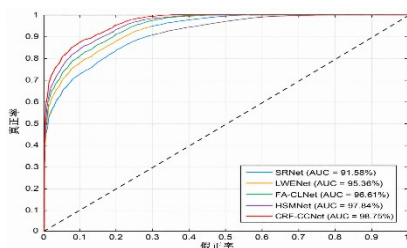


图 7 五种模型在 0.4bpp 下对 WOW 的 ROC 曲线

Fig. 7 ROC Curves of Five Models against WOW at 0.4 bpp

他模型之上, AUC 值更高, 说明本文方法在不同判别阈值下均保持了较好的检测稳定性。

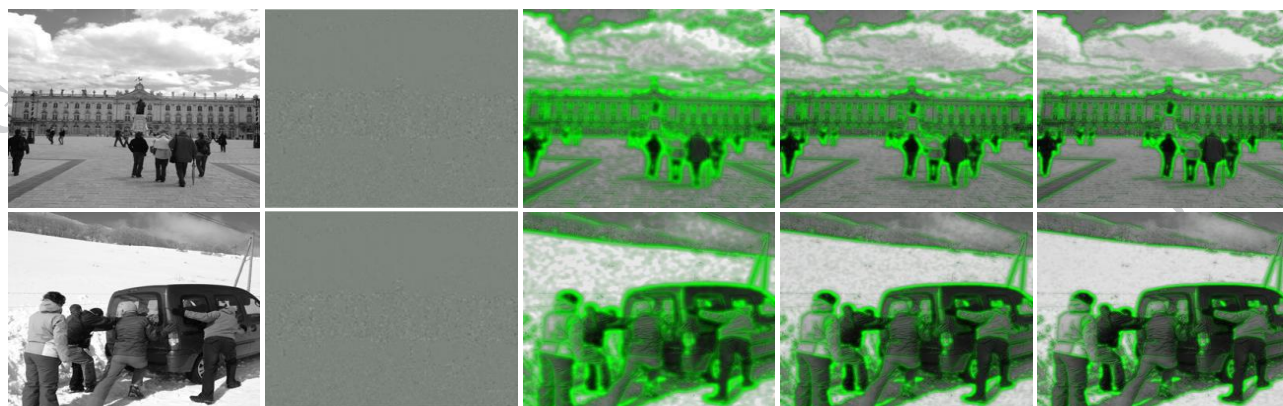
图 8 展示了不同隐写分析模型在 256×256 分辨率图像上的特征响应可视化结果, 其中绿色高亮区域代表模型对隐写敏感特征的响应强度。对比 SRNet、FA-CLNet 以及本文提出的 CRF-CCNet 三种模型的特征响应可以发现: SRNet 与 FA-CLNet 虽能在一定程度上响应图像中的纹理区域与目标边缘, 但其特征图中存在大量无规律的背景噪声响应, 噪声信号与隐写敏感区域的响应强度区分度较低。这一现象反映出传统模型难以有效过滤由平滑背景区

域引入的无关信息, 导致模型在学习过程中易受非隐写噪声干扰, 削弱了对微弱隐写痕迹的聚焦能力。

相比之下, 本文提出的 CRF-CCNet 生成的特征响应更集中: 模型能够精准聚焦于图像边缘、纹理细节等高隐写风险区域, 同时有效抑制了背景区域的无关噪声响应, 特征响应的“信噪比”得到明显提升。这种对隐写敏感特征的精准捕捉能力, 得益于模型中多路径特征增强模块与对比聚类损失的协同作用, 使网络在学习过程中更关注隐写扰动引起的局部结构变化, 而非全局背景信息。这一可视化结果直观证明了 CRF-CCNet 在特征提取阶段对隐写信号的响应更具针对性, 为复杂场景下的隐写分析提供了更可靠的特征基础。

2.5 消融实验

在本节中, 主要讨论模型架构中一些组件对性能的影响, 并通过一系列消融实验验证设计的各个模块的合理性。在所有实验中, 我们都对 BOSS 和 BOSS+BOWS2 数据集进行了比较。为系统验证各创新模块对整体性能的增益, 构建了递进式的模块级消融实验。完整模型可表示为:



(a) 隐写图像 (b) 隐藏的信息 (c) SRNet (d) FA-CLNet (e) CRF-CCNet
(a)stego-Image; (b)Hidden-message; (c)SRNet; (d)FA-CLNet; (e)CRF-CCNet

图 8 部分模型在 256×256 分辨率下的可视化图

Fig. 8 Visualization of Partial Models at 256×256 Resolution

$M_{full} = \text{MultiBand} + \text{MP} - \text{FEM} + \text{MS} - \text{CCL}$ (12)

式中, MultiBand 代表等分辨率多频带残差融合预处理层, MP - FEM 代表多路径特征增强模块, MS - CCL 代表多尺度对比学习损失。为验证各组成模块对隐写分析性能的贡献, 本文在 WOW 隐写算法 0.4 bpp 嵌入率下进行了消融实验: (A) 不引入任何创新模块, 选择 FA-CLNet 隐写分析模型作为基线模型; (B) 在基线模型的基础上, 将预处理层替换为本文提出的等分辨率频带残差建模, 旨在验证预处理模块的有效性。 (C) 在基线模型的基础上, 将原有特征提取模块替换为本文提出的多路径特征增强模块 (MP-FEM), 用于验证全局上下文特征、局部纹理特征和邻域差分特征的协同建模是否能够提升模型对弱隐写扰动的表征能力。 MP-FEM 由 Swin Transformer 路径、卷积路径和中央差分卷积路径组成, 三条路径分别面向长距离上下文依赖、局部纹理模式和边缘残差响应进行特征提取。为进一步分析各路径对检测性能的贡献, 并验证门控融合机制相较于简单拼接的有效性, 本文在保持其他训练设置一致的前提下, 对 MP-FEM 内部结构进行了补充消融实验, 结果如表 2 所示。该实验以基线模型为基础, 仅替换特征提取模块, 不引入 MultiBand 预处理模块和 MS-CCL 损失, 以避免其他模块对 MP-FEM 内部结构分析造成干扰。 (D) 在基线模型的基础上, 仅将其分类阶段替换为研究提出的多尺度对比聚类损失 (MS - CCL)。为了进一步验证所提出 MS-CCL 对特征空间分布优化的有效性, 研究采用 t-分布随机邻域嵌入 (t-distributed Stochastic Neighbor Embedding,

t-SNE) 对模型最终嵌入特征进行了二维可视化分析, 如图 9 所示。该结果进一步证明了研究所提出多尺度对比聚类机制能够有效增强隐写特征的判别表达能力, 从而提升模型对复杂隐写痕迹的检测性能。 (E) 在模型 B 的基础上, 进一步将原有的特征提取模块替换为研究设计的多路径特征增强模块, 验证等分辨率频带残差建模与多路径特征增强模块 (MP-FEM) 的共同效果。 (F) 在 E 基础上, 最终引入 MS - CCL 多尺度对比学习损失, 构成完整 CRF-CCNet, 用于验证模型整体对于隐写检测效果。

整体模型性能的消融实验结果如表 3 所示。 Model A 为基线模型, 其检测准确率为 93.05%。在基线模型基础上引入等分辨率多频带模块后, Model B 的检测准确率提升至 93.72%, 较基线模型提高 0.67%, 表明多频带残差增强能够从不同频率范围中提取更加丰富的隐写响应信息, 有助于增强模型对弱隐写噪声的感知能力。仅引入多路径特征增强模块 (MP-FEM) 时, Model C 的检测准确率达到 94.14%, 相比基线模型提高 1.09%, 说明全局上下文特征、局部纹理特征以及边缘差分特征的协同建模能够有效提高隐写特征表达能力。仅引入多尺度对比聚类损失 (MS-CCL) 时, Model D 的检测准确率提升至 93.57%, 较基线模型提高 0.52%, 表明该损失函数能够在一定程度上优化载体图像与隐写图像在特征空间中的分布, 增强两类样本的可分性。

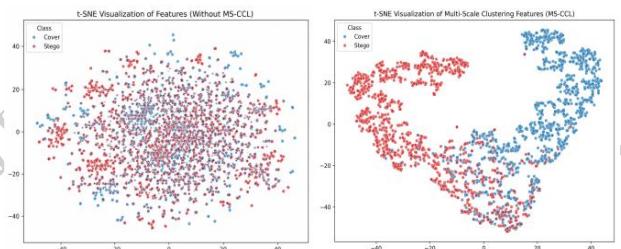
当同时引入 Multi-Band 和 MP-FEM 时, Model E 的检测准确率进一步提升至 94.40%, 相比基线模型提高 1.35%, 并且高于单独引入 Multi-Band 的

Model B 和单独引入 MP-FEM 的 Model C, 说明等分辨率多频带残差增强与多路径特征学习之间具有较好的互补性。进一步在 Model E 的基础上加入 MS-CCL 后, 最终模型 CRF-CCNet(Ours F) 的检测准确率达到 94.81%, 较基线模型提高 1.76%, 较 Model E 进一步提高 0.41%。该结果表明, 多尺度对比聚类约束能够进一步压缩类内特征距离并扩大类间判别间隔, 从而提高模型对微弱隐写痕迹的区分能力。

综合来看, 三个模块均对模型性能提升具有积极作用。其中, MP-FEM 带来的单独提升最为明显, 说明多路径特征建模对隐写分析性能提升具有重要作用; Multi-Band 从频带残差层面补充了不同尺度的隐写信息; MS-CCL 则从特征空间优化角度进一步增强了模型的分类判别能力。三者协同作用使 CRF-CCNet 在消融实验中取得最高检测准确率, 验证了本文所提出网络结构设计的合理性和有效性。

表 2 MP-FEM 内部结构消融实验
Table 2 Ablation Experiment of the Internal Structure of MP-FEM

MP-FEM 模块	Transformer 路径	CNN 路径	CDC 路径	门控 融合	ACC (%)
单 CNN 路径		✓			93.13
CNN+Transformer	✓	✓			93.47
CNN+CDC		✓	✓		93.72
三路径简单拼接	✓	✓	✓		93.95
三路径门控融合	✓	✓	✓	✓	94.14



(a)无 MS-CCL (b)有 MS-CCL
(a) none (b)MS-CCL

图 9 有无多尺度对比聚类损失下特征分布 t-SNE 可视化
Fig. 9 t-SNE Visualization of Feature Distributions With and Without MS-CCL

2.6 性能分析

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

表 3 消融实验
Table 3 Ablation Experiment

模型模块 (Model)	基线 模型	Multi- Band	MP- FEM	MS- CCL	Accuracy -w/o(%)
BaselineA	✓				93.05%
Model B	✓	✓			93.72%
Model C	✓		✓		94.14%
Model D	✓			✓	93.57%
Model E	✓	✓	✓		94.40%
Ours F	✓	✓	✓	✓	94.81%

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。

由表 4 可以看出, 当训练数据由 BOSS 扩展为 BOSS+BOWS2 后, 各隐写分析模型的误检率整体呈下降趋势, 说明增加训练样本能够使模型学习到更充分的图像内容分布和隐写残差特征。以 0.2 bpp 为例, SRNet 的误检率由 0.2961 降至 0.2089, LWENet 由 0.2433 降至 0.1422, FA-CLNet 由 0.2231 降至 0.1312, HSMNet 由 0.2265 降至 0.1400, 本文提出的 CRF-CCNet 则由 0.1849 降至 0.1109。可以看出, 数据规模扩大后, 所有模型均获得了一定性能提升, 但 CRF-CCNet 始终保持最低误检率。进一步在 0.4 bpp 条件下比较, CRF-CCNet 的误检率由 BOSS 数据集上的 0.1019 降至 BOSS+BOWS2 数据集上的 0.0519, 绝对下降 0.0500, 相对降幅约为 49.1%, 表明该模型在较大规模训练数据下能够更充分地发挥特征学习能力。

从不同模型的横向比较来看, CRF-CCNet 在两种数据集和两种嵌入率条件下均取得最低误检率。在仅使用 BOSS 数据集时, CRF-CCNet 在 0.2 bpp 和 0.4 bpp 下的误检率分别为 0.1849 和 0.1019, 均低于其他对比方法。在 0.2 bpp 条件下, 相比次优模型 FA-CLNet 的 0.2231, 误检率降低 0.0382; 在 0.4 bpp 条件下, 相比次优模型 HSMNet 的 0.1102, 误检率降低 0.0083。在 BOSS+BOWS2 数据集上, CRF-CCNet 在 0.2 bpp 和 0.4 bpp 下的误检率分别为 0.1109 和 0.0519。0.2 bpp 条件下相比次优模型 FA-CLNet 的 0.1312 降低 0.0203; 0.4 bpp 条件下相比次优模型 HSMNet 的 0.0602 降低 0.0083。上述结果说明, CRF-CCNet 并非只在单一数据集或单一嵌入率下有效, 而是在不同训练数据规模和不

同嵌入强度下均表现出较好的检测稳定性。

表 4 两种数据集下不同算法的隐写分析误检率比较
Table 4 Comparison of Steganalysis Error Rates for Different Algorithms on Two Datasets

算法	BOSS		BOSS + BOWS2	
	0.2 bpp	0.4 bpp	0.2 bpp	0.4 bpp
SRNet	0.2961	0.1459	0.2089	0.1313
LWENet	0.2433	0.1441	0.1422	0.0941
FA-CLNet	0.2231	0.1367	0.1312	0.0795
HSMNet	0.2265	0.1102	0.1400	0.0602
CRF-CCNet	0.1849	0.1019	0.1109	0.0519

为进一步验证所提出模型在性能与效率之间的平衡能力,本文又将 CRF-CCNet 与经典隐写分析模型 SRNet 以及轻量级模型 LWENet 从模型复杂度进行了对比,结果如表 5 所示。

从模型规模来看,SRNet 参数量达到 4.77 M,而 LWENet 和本文方法分别为 0.82 M 和 0.67 M。相比 LWENet,CRF-CCNet 参数量减少了 0.15 M,降幅约 18.3%;相比 SRNet 则减少约 85.9%,说明所提出网络在保持较强特征表达能力的同时有效降低了模型存储开销。

在计算复杂度方面,SRNet 的 FLOPs 达到 114.20 G,而 LWENet 和 CRF-CCNet 分别为 19.45 G 和 20.25 G。虽然本文方法相较于 LWENet 增加了 0.80 G FLOPs,但仅增长约 4.1%,整体仍保持在同一计算量级,远低于 SRNet,表明所提出的多频带残差建模和多路径特征增强结构并未增加额外计算负担。

在推理效率方面,LWENet 在 RTX 3090 平台上的推理速度为 85 FPS,CRF-CCNet 达到 82.4 FPS,两者差异仅为 2.6 FPS,均能够满足实时检测需求;而 SRNet 由于网络结构较深,其推理速度仅为 18.5 FPS,明显低于轻量级模型。结果表明,本文方法在保持较高检测性能的同时仍具备较好的运行效率。

从检测精度来看,CRF-CCNet 取得了 94.81% 的准确率,显著优于 SRNet 的 86.87% 和 LWENet 的 90.59%。相比 SRNet 提高了 7.94 个百分点,相比 LWENet 提高了 4.22 个百分点。在参数量减少 18.3% 的情况下仍获得更高的检测精度,说明所提出的等分辨率多频带建模、多路径特征增强以及多

尺度对比聚类策略能够有效提升隐写噪声表征能力和特征判别能力。

表 5 性能分析
Table 5 Performance Analysis

Method	Params(M)	FLOPs (G)	FPS	ACC(%)
SRNet	4.77	114.20	18.5	86.87%
LWENet	0.82	19.45	85	90.59%
Ours	0.67	20.25	82.4	94.81%

3 结 论

本文提出了一种针对图像隐写分析任务的多尺度特征融合网络 CRF-CCNet。该网络在预处理阶段通过等分辨率多频带残差提取增强不同频带中的隐写信号响应,在特征提取阶段通过多路径特征增强模块(MP-FEM)联合建模全局上下文信息、局部纹理特征以及边缘噪声变化,并在分类阶段引入多尺度对比聚类损失(MS-CCL)以提升特征判别能力。通过实验验证,CRF-CCNet 在 BOSSBase1.01、BOWS2 等多个主流隐写数据集上,针对 WOW、HILL、S-UNIWARD 等隐写算法及不同嵌入率条件下均表现出优异的检测性能。同时,该网络在保持参数量低于 LWENet 的基础上,实现了比 SRNet 和 LWENet 更高的检测准确率,显示了在性能与计算开销之间的良好平衡。

本研究的主要贡献体现在以下三个方面:

频带建模创新:通过等分辨率多频带残差预处理,使网络能够同时捕获高频、中频及低频隐写特征,增强了对不同频带嵌入扰动的响应能力。多路径特征融合:MP-FEM 结合 Transformer 全局依赖建模、深度可分离卷积局部纹理提取及中央差分卷积边缘增强,实现隐写噪声的多维联合表征,提高了特征表达能力。多尺度对比聚类优化:MS-CCL 通过动态原型约束与跨尺度一致性约束,有效压缩类内分布并扩大类间间隔,提升了模型对弱隐写样本的判别能力和泛化性能。

CRF-CCNet 在检测性能、模型轻量化和实际应用价值方面均表现出一定优势,可为数字版权保护、信息安全监测及图像取证等任务提供技术支撑。同

时,本文方法仍存在一定局限。首先,当前实验主要在灰度图像上进行验证,模型在彩色图像载体上的检测能力仍需进一步研究。其次,MS-CCL和MP-FEM模块依赖充分标注的训练数据,在低标注或无标注条件下的性能可能受到影响。最后,本文方法主要面向修改式隐写检测,其判别依据是载体图像经嵌入修改后产生的微弱残差扰动,因此并不直接适用于完全生成式隐写检测。生成式隐写通常不再以同一载体图像为基础进行像素级修改,而是通过生成模型直接生成含密图像,其检测依据更可能表现为生成分布偏差、纹理统计异常或模型指纹,而非传统意义上的嵌入残差。因此,本文的频带残差建模和对比聚类思想可为生成式隐写检测提供一定参考,但仍需结合生成式图像的特征重新设计训练数据、判别目标和特征约束机制。

未来工作方向主要包括:探索自监督或半监督学习方法,以提升在低标注条件下的特征建模能力;将CRF-CCNet框架扩展至彩色图像及AIGC生成图像,提升在多样化载体上的判别能力。探索面向生成式隐写的判别机制,并研究本文多频带残差建模与多尺度对比学习策略在生成式隐写场景中的可迁移性。

综上所述,CRF-CCNet提供了一种有效的多尺度特征融合方案,通过频带增强、多路径建模和对比聚类优化,实现了高效、准确且轻量的图像隐写分析,为隐写分析领域提供了新的技术思路和实践依据。

参考文献(References)

Boroumand M, Chen M and Fridrich J. 2019. Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 14(5): 1181-1193 [DOI: 10.1109/TIFS.2018.2871749]

Chen J F, Fu Z J, Zhang W M, Cheng X and Sun X M. 2021. A survey of image steganalysis based on deep learning. *Journal of Software*, 32(2): 551-578 (陈君夫,付章杰,张卫明,程旭,孙星明. 2021. 基于深度学习的图像隐写分析综述. *软件学报*, 32(2): 551-578) [DOI: 10.13328/j.cnki.jos.006135]

Chen Z Y, Yin Z X, Zhan H J, Lyu S J and Hu M H. 2025. Reversible data hiding in encrypted images based on combined encoding method containing the opposite bit. *Journal of Image and Graphics*, 30(3): 769-783 (陈振宇,殷赵霞,占鸿渐,吕淑静,胡孟晗. 2025. 隐含异位联合编码的密文图像可逆信息隐藏. *中国图象*

图形学报, 30(3): 769-783) [DOI: 10.11834/jig.240287]

Fu T, Chen L Q, Jiang Y H, Zhang Y Q, Li X W, Wang J H, et al. 2025. Image steganalysis based on dual-path enhancement and fractal downsampling. *IEEE Transactions on Information Forensics and Security*, 20: 1-16 [DOI: 10.1109/TIFS.2024.3493615]

Fridrich J and Kodovský J. 2012. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3): 868-882 [DOI: 10.1109/TIFS.2012.2190402]

He J, Weng S W, Yu L F, Zhang Y Q, Chen M and Sun S. 2025. Steganalysis network with two-branch preprocessing for spatial and JPEG domains. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(2): 1451-1463 [DOI: 10.1109/TCSVT.2024.3470809]

He Z X, Wu R L and Wang X Y. 2025. Image steganalysis based on an adaptive attention mechanism and lightweight DenseNet. *Computers, Materials & Continua*, 85(1): 1631-1651 [DOI: 10.32604/cmc.2025.067252]

Holub V, Fridrich J and Denemark T. 2014. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security*, 2014(4): 1-13 [DOI: 10.1186/1687-417X-2014-1]

Holub V and Fridrich J. 2012. Designing steganographic distortion using directional filters//*Proceedings of the IEEE International Workshop on Information Forensics and Security*. Tenerife, Spain: IEEE: 234-239 [DOI: 10.1109/WIFS.2012.6412655]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, United States: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, United States: IEEE: 4700-4708 [DOI: 10.1109/CVPR.2017.243].

Jia M L, Yang Y and Sun D. 2024. Reversible image hiding with super-resolution behavior camouflage effect. *Journal of Image and Graphics*, 29(2): 382-394 (贾孟霖,杨杨,孙冬. 2024. 具有超分辨率行为伪装效果的可逆图像隐藏. *中国图象图形学报*, 29(2): 382-394) [DOI: 10.11834/jig.230465]

Kodovský J, Fridrich J and Holub V. 2012. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security*, 7(2): 432-444 [DOI: 10.1109/TIFS.2011.2175919]

Li B, Wang M, Huang J and Li X. 2014. A new cost function for spatial image steganography//*Proceedings of the IEEE International Conference on Image Processing*. Paris, France: IEEE: 4206-4210 [DOI: 10.1109/ICIP.2014.7025854]

Li S Y, Han Y F, Le Y F, Yao H and Qin C. 2024. Constructive information hiding with texture generation driven by B-spline. *Journal of*

- Image and Graphics, 29(4): 989-1002 (酃姝伊, 韩彦芳, 乐燕芬, 姚恒, 秦川. 2024. B样条驱动纹理生成的构造式信息隐藏. 中国图象图形学报, 29(4): 989-1002) [DOI: 10.11834/jig.221173]
- Li W Q, Zhou K Y, Hu X X, Zhang X P, Li S and Qian Z X. 2025. Generative video steganography via secret mapping in diffusion models. Journal of Image and Graphics, 30(11): 3465-3478 (李文琪, 周科扬, 胡宵宵, 张新鹏, 李晟, 钱振兴. 2025. 扩散模型中秘密信息映射的生成式视频隐写. 中国图象图形学报, 30(11): 3465-3478) [DOI: 10.11834/jig.250079]
- Long L H, Wang Z C and Zhang X P. 2026. Overview of neural network model steganography. Journal of Image and Graphics, 31(1): 45-61 (龙玲慧, 王子驰, 张新鹏. 2026. 神经网络模型隐写研究进展. 中国图象图形学报, 31(1): 45-61) [DOI: 10.11834/jig.250267]
- Luo G, Wei P, Zhu S, Zhang X, Qian Z and Li S. 2022. Image steganalysis with convolutional vision transformer//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore, Singapore; IEEE: 3089-3093 [DOI: 10.1109/ICASSP43922.2022.9747091]
- Luo W W, Liu C L and Lei Q. 2025. Spatial image steganalysis based on preprocessing layer enhancement and attention mechanism. Journal of Jilin University (Engineering and Technology Edition), 55(12): 4024-4033 (罗维薇, 刘长龙, 雷琴. 2025. 基于预处理层增强和注意力机制的空域图像隐写分析. 吉林大学学报(工学版), 55(12): 4024-4033) [DOI: 10.13229/j.cnki.jdxbgxb.20240437]
- Ma Y Y, Xu L G, Zhang Q Q, Zhang Y, Xin X W and Luo X Y. 2025. EIS-OBEA: enhanced image steganalysis via opposition-based evolutionary algorithm. IEEE Transactions on Information Forensics and Security, 20: 3616-3631 [DOI: 10.1109/TIFS.2025.3549692]
- Ma B, Li K, Xu J, Wang C P, Li J and Zhang L W. 2024. High-security image steganography with the combination of multiple competition and channel attention. Journal of Image and Graphics, 29(2): 355-368 (马宾, 李坤, 徐健, 王春鹏, 李健, 张立伟. 2024. 联合多重对抗与通道注意力的高安全性图像隐写. 中国图象图形学报, 29(2): 355-368) [DOI: 10.11834/jig.230134]
- Pevný T, Bas P and Fridrich J. 2010. Steganalysis by subtractive pixel adjacency matrix. IEEE Transactions on Information Forensics and Security, 5(2): 215-224 [DOI: 10.1109/TIFS.2010.2045842]
- Peng C, Zhang X, Yu G, Luo G and Sun J. 2019. Dual attention network for scene segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(5): 1745-1758 [DOI: 10.1109/TPAMI.2019.2946129]
- Suykens J A K and Vandewalle J. 1999. Least squares support vector machine classifiers. Neural Processing Letters, 9: 293-300 [DOI: 10.1023/A:1018628609742]
- Tan M and Le Q V. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, United States: PMLR: 6105-6114.
- Weng S, Chen M, Yu L and Sun S. 2022. Lightweight and effective deep image steganalysis network. IEEE Signal Processing Letters, 29: 1888 - 1892 [DOI: 10.1109/LSP.2022.3201727]
- Xu G, Wu H Z and Shi Y Q. 2016. Structural design of convolutional neural networks for steganalysis. IEEE Signal Processing Letters, 23(5): 708-712 [DOI: 10.1109/LSP.2016.2548421]
- Yan L, Li Z, Liu J, Dong Y and Zhang J. 2025. Feature aware-contrastive learning network for arbitrary-sized image steganalysis. Journal of Visual Communication and Image Representation, 111: 104525 [DOI: 10.1016/j.jvcir.2025.104525]
- Yang W, Liu Y, Chen Y, Cui Y, Guo C, Shen G, et al. 2026. HSMNet: a multi-resolution grayscale image steganalysis method based on hybrid dilated convolution and self-attention multi-channel network. Information Sciences, 728: 122824 [DOI: 10.1016/j.ins.2025.122824].
- Ye J, Ni J and Yi Y. 2017. Deep learning hierarchical representations for image steganalysis. IEEE Transactions on Information Forensics and Security, 12(11): 2545-2557 [DOI: 10.1109/TIFS.2017.271094]

作者简介

刘俊彤, 女, 硕士研究生, 主要研究方向为信息安全, 隐写分析。E-mail: 12242019@stu.lzjtu.edu.cn

甘博宇, 男, 硕士研究生, 主要研究方向为信息安全, 隐写分析。E-mail: 11240707@stu.lzjtu.edu.cn