# A Lightweight Steganographic Scheme for Color Images

**Chen Liu**

School of Cyberspace Security, Jiangsu Police Institute, Nanjing, Jiangsu, 210031, China

**Abstract**

To address the shortcomings of existing deep learning-based image steganography methods, such as numerous modules and high training overhead, a lightweight steganographic scheme for color images in the wavelet domain is proposed. In this scheme, the features of the cover image are obtained based on wavelet transform. The encoded stego image is concatenated with the low-frequency subband of the cover image to generate a perturbation, which is then embedded into the high-frequency subbands of the cover image. Finally, the stego cover image is obtained through inverse wavelet transform. Experiments on the COCO val2017 dataset show that when the payload capacity is 24 bpp, the PSNR of the stego cover image is 33.36 dB, and the PSNR of the recovered stego image is 31.44 dB.

**Keywords**

information steganography; wavelet transform; deep learning; lightweight network

# 面向彩色图像的轻量级隐写方案

刘琛

江苏警官学院网络空间安全学院，中国·江苏 南京 210031

**摘要**

针对现有深度学习图像隐写方法模块繁多、训练开销大的不足，提出一种轻量级小波域彩色图像隐写方案。该方案基于小波变换获得载体图像的特征，编码后的隐写图像与载体图像的低频子带拼接生成扰动，嵌入到载体图像的高频子带后，再经逆小波变换得到含密载体图像。在COCO 2017数据集上的实验表明，当载密率为24bpp时，含密载体图像PSNR为33.36dB，恢复隐写图像PSNR为31.44dB。

**关键词**

信息隐写；小波变换；深度学习；轻量级网络

## 1 引言

通过让模型自主学习嵌入提取策略，基于深度学习的隐写方法在载体及含密信息的表征能力上显著增强。为提升性能，研究者一方面借助可逆神经网络构造严格可逆的变换，使载体能够在提取端无损还原[1]。另一方面，借助生成式模型实现含密载体图像的高保真度还原[2-3]。这些方法性能优异，结构一般比较复杂，对算力有较高要求。本文面向轻量级应用场景，提出一种结构简单的小波域残差嵌入彩色图像隐写方案。载体图像经小波变换分解为低频与高频子带，隐写图像编码后与低频子带拼接，以学习载体的先验信息，使不可感知性由载体本身提供。拼接后经扰动网络产生的扰动项叠加于高频子带，通过纹理边缘区域对修改不敏感的特性，从而降低对载体的影响。实验表明，该方案可训练参数较小，在单卡环境下仅数个小时就可以完成训练，具有很强的实用性。

【基金项目】江苏省高校哲学社会科学研究一般项目（项目编号：2023SJYB0463）。

【作者简介】刘琛（1990-），男，中国江苏南京人，博士，副教授，从事信息隐写、计算机视觉研究。

## 2 隐写方法

### 2.1 整体框架

图1给出了基于小波变换与深度神经网络的图像隐写与提取框架。在嵌入阶段，首先对载体图像进行小波变换，将其分解至频域以获取多尺度特征。同时，隐写图像经由隐写编码器进行特征编码。随后，编码后的隐写特征与载体图像的小波域特征共同输入至扰动生成网络，生成用于嵌入的扰动信息。该扰动信息通过残差嵌入模块被注入到载体图像的小波系数中，再经由逆小波变换重构回空域，从而生成视觉上无明显异常的含密载体图像。在提取阶段，接收端对含密载体图像再次执行小波变换，将其转换回频域，随后利用隐写解码器从频域系数中解析并恢复出原始的隐写图像。该架构利用频域变换的特性，旨在提升隐写容量与不可感知性之间的平衡，并确保信息提取的鲁棒性。

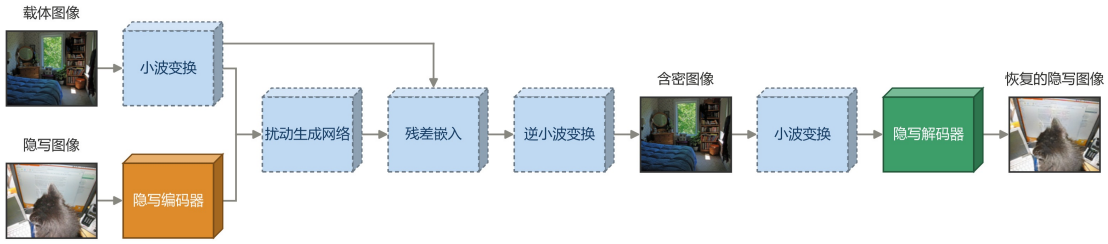



图1 基于小波变换的轻量级图像隐写框架

### 2.2 隐写模块

隐写模块包含隐写编码器、小波变换、扰动生成网络、残差嵌入以及逆小波变换模块。尺寸为3×128×128的彩色隐写图像S经隐写编码器被转化为9×64×64的信息。其作用是把空间域的隐写图像变换到与高频子带同维度的特征量，使后续扰动生成网络可以在同一尺度上直接使用。

$$\Delta = \alpha \cdot \tanh(P([LL, S_h])) \quad (1)$$

$$HF' = HF + \Delta \quad (2)$$

如式（1）和（2）所示，编码后的隐写图像信息（S_h）与载体图像的低频子带信息（LL）进行拼接。经过扰动生成网络产生扰动，之后嵌入到载体图像的三个高频子带中。式中$P(\cdot)$表示扰动生成网络，$[,]$表示通道维拼接，$\alpha$为可学习的幅度系数，tanh为双曲正切函数，目的是把扰动信息转化到$(-1,1)$区间，避免个别位置出现过大失真。由于扰动来源包含载体的低频子带信息，模型可以学习到载体的特征，从而保证含密载体的不可感知性。此外，高频子带维度为9×64×64，可以承载与载体尺寸相同的彩色隐写图像，从而有效提升了载密率。最后，残差嵌入的信息经过逆小波变换产生含密载体图像。

### 2.3 提取模块

提取模块由小波变换和隐写解码器组成。含密载体图像经小波分解后，得到含密载体图像的低频子带LL'和高频子带HF'。隐写解码器同时使用低频与高频两路子带作为输入，其中高频子带携带隐写信息，低频子带提供场景上下文，有助于稳定恢复。

### 2.4 损失函数

训练过程的总损失由隐写损失L_stego和恢复损失L_secret两部分组成，如式（3）所示。

$$L = L_stego + L_secret$$

$$L_stego = \beta \cdot MSE(C', C) + (1-\beta) \cdot (1 - SSIM(C', C))$$

$$L_secret = \beta \cdot MSE(S', S) + (1-\beta) \cdot (1 - SSIM(S', S)) \quad (3)$$

L_stego表示载体图像嵌入前后的损失，L_secret表示隐写图像恢复前后的损失。两项均由均方误差与结构相似度加权组成，$\beta$为MSE项权重。

## 3 实验结果与分析

### 3.1 实验环境与数据集

实验硬件为单张NVIDIA RTX 3090，软件环境为Python 3.10、CUDA 12.6。训练与验证均使用COCO 2017数据集，图像统一缩放至128×128，采用RGB三通道。隐写图像与载体图像同尺寸，载密率24bpp，二者随机配对。批量大小为16，单次训练50轮。

### 3.2 隐写与提取效果

图2给出了一组实验结果。从左至右依次为载体图像、含密载体图像、含密载体图像与载体图像的逐像素归一化残差值、原始隐写图像以及恢复的隐写图像。可以看到，含密载体图像与载体图像在视觉上具有高度一致性，含密载体图像与载体图像的逐像素归一化残差值图仅在物体的轮廓与纹理处呈现微弱响应，平坦背景区域几乎无差异。这表明隐写模块将信息融入到载体图像的同时，保证了不可感知性的要求。恢复的隐写图像完整保留了原图的结构与配色，无明显的失真，验证了信息提取的准确性。

### 3.3 隐写性能指标

图3给出了训练过程曲线。从训练损失曲线中可以看出，隐写损失L_stego和恢复损失L_secret在前5轮内迅速下降后趋于平稳，表明训练过程稳定、未出现发散。在恢复质量指标曲线中，含密载体的PSNR在前5轮内由约15dB升至32dB以上，随后维持在32-34dB区间。恢复隐写图像PSNR呈类似的变化，在训练结束时提升到32dB左右。从扰动幅度系数$\alpha$与高频干扰比变化曲线中可以发现，$\alpha$在训练早期短暂上升后主动下降并稳定在0.6附近，$|\Delta|/|HF|$相应由1.2降至0.4左右。这一现象说明模型逐渐收敛，在嵌入相同信息量的情况下产生的扰动不断变小。

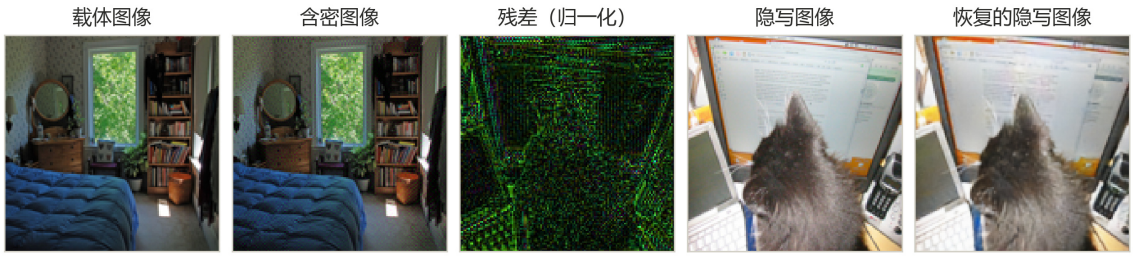



图 2 隐写与提取效果

最后，基于验证集的实验给出了本方案的隐写性能指标，如表1所示。在24bpp这一较高载密率下，含密载体图像PSNR为33.36dB，视觉上已难以与载体区分。恢复的隐写图像PSNR为31.44dB、SSIM为0.9653，结构与配色均得到良好保持。

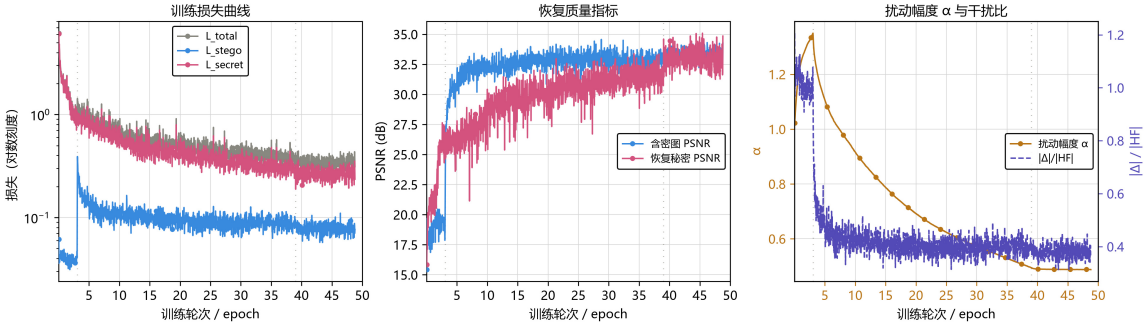



图 3 训练过程曲线

表 1 隐写性能指标

| 指标 | PSNR/dB | SSIM |
|---|---|---|
| 含密载体图像 | 33.36 | 0.9411 |
| 恢复隐写图像 | 31.44 | 0.9653 |

## 4 结语

本文提出一种基于小波域残差嵌入的轻量级彩色图像隐写方案。方案以小波变换为基础，保持低频子带不变，仅在高频子带上叠加隐写图像带来的扰动，接收端经隐写提取网络一步输出复原的隐写图像。在COCO 2017数据集上的实验表明，在较高载密率的条件下，含密载体图像在视觉上不可区分、隐写图像可被准确提取。方案模块构成简单，可训练参数较小，对算力要求适中，可在单卡上完成。

## 参考文献

[1] 张春玉,石倚菲,李磊康,等.基于可逆神经网络的多图像隐写方案[J].计算机工程与设计,2025,46(12):3498-3505.

[2] 董云云,张锦弘,李钒效,等.基于双扩散模型的未知源-目标域生成式图像隐写方法[J].计算机研究与发展,2026,63(7):1831-1841.

[3] 宫长威,金波,袁得嵛,等.生成式图像隐写综述：方法、鲁棒性挑战与展望[J].信息网络安全,2026,26(7):1058-1076.