计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(3):180−188 [doi: 10.15888/j.cnki.csa.009798] [CSTR: 32024.14.csa.009798]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 PEMA 和 DWT 的岩心图像鲁棒水印算法①

严宇真[1], 何小海[1], 卿粼波[1], 罗彬彬[2], 滕奇志[1]

[1](四川大学 电子信息学院, 成都 610065)
[2](成都西图科技有限公司, 成都 610065)
通信作者: 何小海, E-mail: hxh@scu.edu.cn

摘　要: 岩心图像作为地质、油气等行业中极为重要的数字图像资源, 对于科学研究和工程实践至关重要, 其安全性常通过添加数字水印的方式来保障. 在数字化的进程中, 岩心图像在存储、传输和网页发布等情况下常会进行 JPEG 压缩. 然而, 现存基于深度学习的图像数字水印算法在应对 JPEG 压缩时, 视觉质量和鲁棒性方面仍存在显著不足. 本文提出了一个端到端图像鲁棒水印算法, 旨在解决 JPEG 压缩条件下岩心图像的鲁棒水印嵌入问题. 为高效融合载体图像与水印的特征, 本文引入了多尺度跨时空注意力 (pyramid efficient multi-scale attention, PEMA) 模块, 该模块通过独特的跨空间交互策略和通道间关系的构建方式, 能够有效捕获不同方向上的长程依赖以及不同尺度下的特征信息. 为了实现视觉不可感知性, 本文通过离散小波变换 (discrete wavelet transform, DWT) 将数字水印嵌入到载体图像的低频分量中, 并引入 DLL (DWT LL sub-band loss) 损失函数, 以提升水印图像的视觉质量. 实验结果表明, 该算法在针对 JPEG 压缩的鲁棒性和视觉不可感知性上均优于现有主流算法.

关键词: 鲁棒水印; JPEG 压缩; 空间多尺度; 注意力机制; 深度学习

引用格式: 严宇真,何小海,卿粼波,罗彬彬,滕奇志.基于 PEMA 和 DWT 的岩心图像鲁棒水印算法.计算机系统应用,2025,34(3):180–188. http://www.c-s-a.org.cn/1003-3254/9798.html

Algorithm of Robust Watermarking for Core Images Based on PEMA and DWT

YAN Yu-Zhen[1], HE Xiao-Hai[1], QING Lin-Bo[1], LUO Bin-Bin[2], TENG Qi-Zhi[1]

[1](College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)
[2](Chengdu Xitu Technology Co. Ltd., Chengdu 610065, China)

Abstract: Core images, as a crucial digital image resource in the fields of geology, oil, and gas, are essential for scientific research and engineering practices. Their security is often ensured by adding digital watermarks. During digitization, core images frequently undergo JPEG compression when they are stored, transmitted, or published on Web pages. However, existing deep learning-based image digital watermarking algorithms still have significant shortcomings in terms of visual quality and robustness under JPEG compression. This study proposes an end-to-end image robust watermarking algorithm to address the issue of robust watermark embedding in core images under JPEG compression conditions. To efficiently integrate the features of the host image and the watermark, the study introduces a pyramid efficient multi-scale attention (PEMA) module. Through a unique cross-spatial interaction strategy and channel-wise relationship construction, the module effectively captures long-range dependencies in different directions and features information at various scales. To achieve visual imperceptibility, the study embeds the digital watermark into the low-frequency components of the host image using discrete wavelet transform (DWT) and introduces the DWT LL sub-band loss (DLL) loss function to improve the visual quality of the watermark image. Experimental results demonstrate that the proposed algorithm outperforms existing mainstream algorithms in both robustness against JPEG compression and visual imperceptibility.

① 基金项目: 国家自然科学基金 (62071315)
收稿时间: 2024-09-11; 修改时间: 2024-10-10; 采用时间: 2024-10-14; csa 在线出版时间: 2025-01-21
CNKI 网络首发时间: 2025-01-22

Key words: robust watermarking; JPEG compression; spatial multi-scale; attention mechanism; deep learning

在当今信息化和数字化的时代, 社交媒体软件已深深植根于人们的日常生活, 成为不可或缺的重要组成部分. 在享受其带来的数据共享与交流的便利的同时, 信息安全和数据保护成为各个领域的重要议题[1]. 数字图像作为广泛应用的多媒体信息之一, 其直观展示的信息在地质勘探和资源开发中尤为重要. 岩心图像作为记录地下地质特征的关键数据, 具有极高的研究和商业价值, 是地质、油气等行业分析和决策的重要依据. 然而, 随着产业信息化的推进, 岩心图像面临篡改、盗用和非法复制等安全威胁, 如何有效保护这些数据成为亟待解决的问题.

数字水印技术自诞生起便广泛应用于版权保护领域, 以实现版权保护、篡改检测和内容认证等多种功能. 该技术将辅助信息嵌入到多媒体介质中, 在保持视觉上的不可感知性的同时实现对多种噪声的鲁棒性[2]. 传统的数字水印算法[3] , 如基于空间域和频域的方法, 虽然取得了一定成果, 但由于其通常围绕特定的算法和模式实现, 在安全性和容量上有明显局限, 且易被研究和仿造, 削弱保护效果. 近年来, 深度学习算法因其出色的特征表示和学习能力, 在图像处理领域崭露头角[4]. 引入深度学习算法的岩心图像数字水印技术, 通过构建模型来学习数据的内在特征和分布, 能够生成更为隐蔽且具鲁棒性的水印, 为岩心图像的保密性提供全新解决方案. 深度学习模型的自适应能力不仅能够抵抗传统攻击手段, 还能应对新型攻击, 提高水印算法的实用性和应用前景. 然而, 大多数基于深度神经网络的数字水印模型在面对 JPEG 压缩时并没有很好的效果[5].

Zhu 等人[6]提出的 HiDDeN 框架利用噪声层对编码器和解码器进行联合训练, 在鲁棒性和图像质量上超越传统算法. Ahmadi 等人[7]的 RedMark 框架使用残差网络和强度因子控制水印强度. Tancik 等人[8]通过二进制编码超链接实现对真实打印和摄影场景扰动的鲁棒性. Fang 等人[9]的 PIMoG 模型通过模拟透视畸变、光照畸变和摩尔畸变, 实现屏幕拍摄的鲁棒性. Liu 等人[10]提出的 TSDL 框架通过分阶段训练消除噪声层限制. Jia 等人[11]的 MBRS 框架通过随机选取不同噪声层, 取得出色效果.

为了在真实 JPEG 压缩场景下更好地实现对于岩心图像的水印嵌入, 本文提出了一个端到端岩心图像鲁棒水印算法. 该算法通过在网络中引入多尺度跨时空注意力模块 (pyramid efficient multi-scale attention, PEMA) 实现对不同尺度的特征捕捉以及载体图像和数字水印的特征融合. 不同于直接将数字水印嵌入到载体图像的空间域中, 本文使用离散小波变换 (discrete wavelet transform, DWT) 对载体图像进行处理, 并将数字水印嵌入到其频域的 LL 分量中, 通过 DLL (DWT LL sub-band loss) 损失函数来实现人眼视觉上的不可感知性.

1 基于 PEMA 和 DWT 的岩心图像鲁棒水印算法

1.1 网络结构

本文的网络结构如图 1 所示, 主要包括 5 个部分: 信息处理器、编码器、对抗判别器、噪声层以及解码器. 本节接下来将对这 5 个部分进行详细介绍, 表 1 展示了本文中所用符号的定义.

1.1.1 信息处理器

信息处理器负责处理要嵌入的水印信息, 将其转换成模型可以理解和处理的格式, 并进行特征提取和处理, 以便与图像特征结合. 信息处理器增加了水印信息的复杂性和安全性, 为编码器提供了额外的保护层. 首先, 长度为 L 的水印信息 M 被重塑为张量 $F_M^{1\times h\times w}$, 其中 $L=h\times w$. 该张量经过一个由 3×3 的卷积、批标准化层以及 ReLU 激活函数 (记为 ConvBNRelu) 组成的卷积模块, 接着被几个步长为 2 的转置卷积模块扩展为形状为 $C\times H\times W$ 的张量后送入到几个 SENet 模块中. 得到的特征图将会送入到编码器中进行进一步的处理.

需要注意的是, 每个转置卷积模块使得输入的特征张量的长和宽翻倍, 因此, 数字水印信息 M 和载体图像 I_{co} 需要满足:

$$L=h\times w=\frac{H}{2^N}\times\frac{W}{2^N} \tag{1}$$

其中, H 和 W 为载体图像的长和宽, N 是由 L、H 和 W 共同决定的整数.

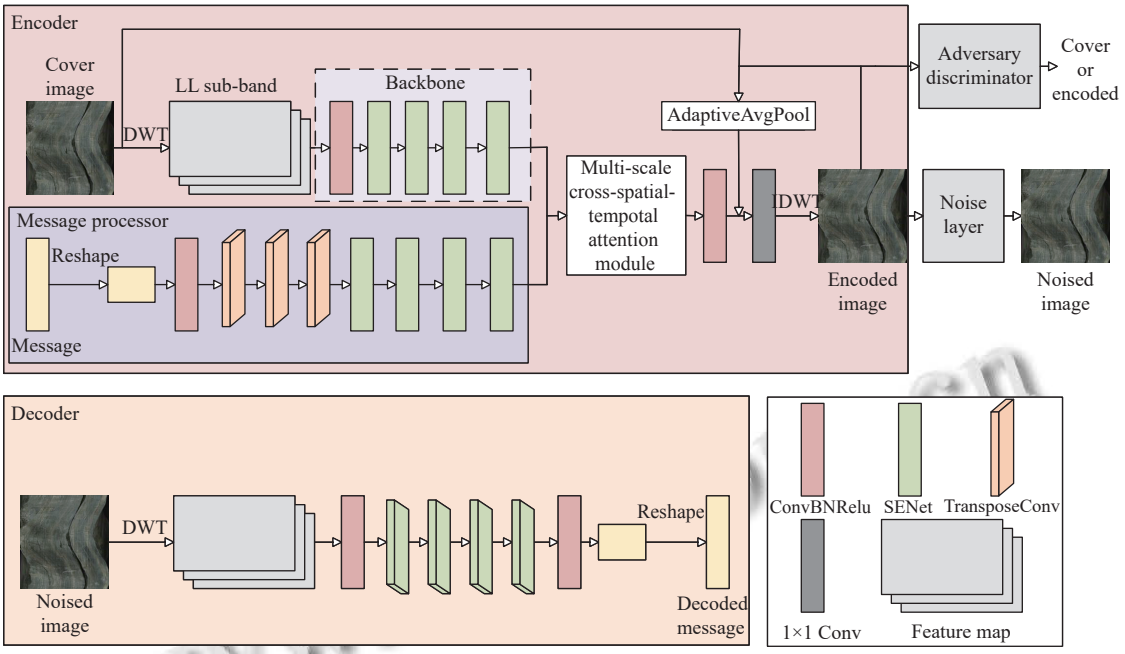

图 1 网络结构

表 1 符号定义

符号	定义
I_{co}	载体图像: 用于嵌入数字水印的图像
I_{en}	编码图像: 已经潜入了数字水印的图像
I_{no}	噪声图像: 经过噪声层后遭到噪声攻击的图像
M	数字水印: 用于嵌入的数字水印信息
M'	复原水印: 从编码图像中复原的数字水印信息

1.1.2 编码器

编码器的主要任务是将水印信息嵌入到岩心图像中. 依赖于深度学习架构强大的特征提取能力, 编码器能够确保水印信息以高度隐秘的方式嵌入到图像的不同区域. 本文的编码器整合了 DWT, 以及 PEMA 模块, 以实现对不同尺度下的特征捕捉以及数字水印在人眼感知下的高不可见性. 载体图像I_{co}首先经过 DWT 变换后被分解为 LL、LH、HL、HH 这 4 个分量, 选取 LL 分量作为数字水印嵌入的载体, 输入 ConvBNRelu 模块当中. 接着, 再输入 4 个 SENet 模块中以提取 LL 分量的特征张量. 将得到的特征张量和信息处理器中得到的特征张量一同输入到 PEMA 中进行进一步的处理, 以进行不同尺度下的特征捕捉与特征融合. 将结果进过 ConvBNRelu 处理后与I_{co}进过平均池化得到的特征进行跳跃连接, 以此来整合低层次的特征信息和高层次的特征信息. 最后经过1×1的卷积以及 IDWT 后得到编码图像I_{en}.

1.1.3 对抗判别器

对抗判别器由核为 3 的卷积和平均池化层构成, 目的是通过训练提升自身识别和区分载体图像I_{co}与嵌入水印后的图像I_{en}之间的细微差异以对抗编码器. 对抗性训练的方法使得编码器尽可能地学习欺骗对抗判别器做出错误判断的方式, 从而提高载体图像I_{en}的质量, 实现人眼感知上的隐蔽性.

1.1.4 噪声层

噪声层为整个网络提供了对抗噪声攻击的鲁棒性. 本文的噪声层包括无噪声、可微分模拟 JPEG 压缩以及真实 JPEG 压缩 3 种, 在训练过程中为每个 Batch 随机选取一种噪声进行攻击. 无噪声层保证了解码器对于没有被 JPEG 压缩攻击的图像的解码能力、真实 JPEG 压缩层保证了解码器对于 JPEG 压缩攻击的鲁棒性、可微分模拟 JPEG 压缩为编码器提供了可以回传的梯度, 使得编码器和解码器可以被联合训练.

1.1.5 解码器

解码器主要目的是从被噪声攻击的图像I_{no}中提取复原后的数字水印信息M', 并尽可能使$M=M'$. 在本文的解码器中, I_{no}先经过 DWT 后被分解为 LL、LH、HL、HH 这 4 个分量. 由于数字水印被嵌入在载体图像I_{co}的 LL 分量中, 故选取 LL 分量送入 ConvBNRelu 中, 之后经过几个 SENet 进行下采样, 将其变为形状为$C\times h\times w$的张量. 最后通过 ConvBNRelu 将其压缩至单

通道, 并进行重塑以得到解码信息M'.

1.2 多尺度跨时空注意力模块

1.2.1 高效多尺度注意力机制 (EMA)

在轻量级网络上的研究表明, 通道注意力会给模型带来比较显著的性能提升, 但是通道注意力通常会忽略对生成空间选择性注意力图非常重要的位置信息. 因此, CBAM[12]注意力模块应运而生. 其通过对通道注意力和空间注意力两个模块的结合, 使得网络可以同时学习通道和空间上的信息, 但这也为其带来了计算复杂度较高的缺点.

此后, Hou 等人[13]提出了一种为轻量级网络设计的高效注意力机制. 其通过在通道特征中融入特征图的空间特征, 进而有效地提取出一个方向的长程依赖, 进而生成相应的注意力特征. 然而, CA 注意力分别捕捉了 H 和 W 方向上的全局依赖性后, 仅通过一个核为 1 的卷积融合通道特征表示, 并没有考虑到不同尺度下的空间信息, 且通过通道降维的方式来建模通道间的关系会对提取深度视觉表征带来副作用. 因此, Ouyang 等人[14]提出了一种基于跨空间学习的高效多尺度注意力模块 (efficient multi-scale attention module with cross-spatial learning), 将部分通道重塑为 batch 维度, 并分组为多个子特征, 使得空间语义特征均匀分布在每个特征组内. 同时, 两条并行分支的输出特征通过矩阵乘法的方式融合, 以实现不同尺度空间下的信息聚合. EMA 的结构如图 2 所示. EMA 有两条并行的分支, 其中一条用类似 CA 的方法, 在通道特征中融入某一个方向上的长程依赖, 不同的是, 其并没有像 CA 一样用通道降维的方式来建立通道间的关系而是仅利用 Concat 和核为 1 的卷积来进行特征拼接与融合通道特征表示. 之后分割后又分别经过 Sigmoid 函数后生成特征权重表示, 并以此来分配权重. 由此得到的是 1×1 尺度的特征. 另一条分支通过核为 3 的卷积来捕获不同尺度下的特征信息. 可以看到, 1×1 的空间特征经过正则化、平均池化和 Softmax 函数之后生成 1×1 尺度的通道描述符. 将该通道描述符与 3×3 尺度下的特征进行矩阵乘法得到 1×1 尺度权重下的 3×3 尺度全局空间注意力表示. 同理, 另一分支得到了 3×3 尺度权重下的 1×1 尺度全局空间注意力表示. 将这两个不同尺度的空间注意力相加以得到跨空间聚合信息, 最后通过 Sigmoid 函数生成权重来重新调整输入.

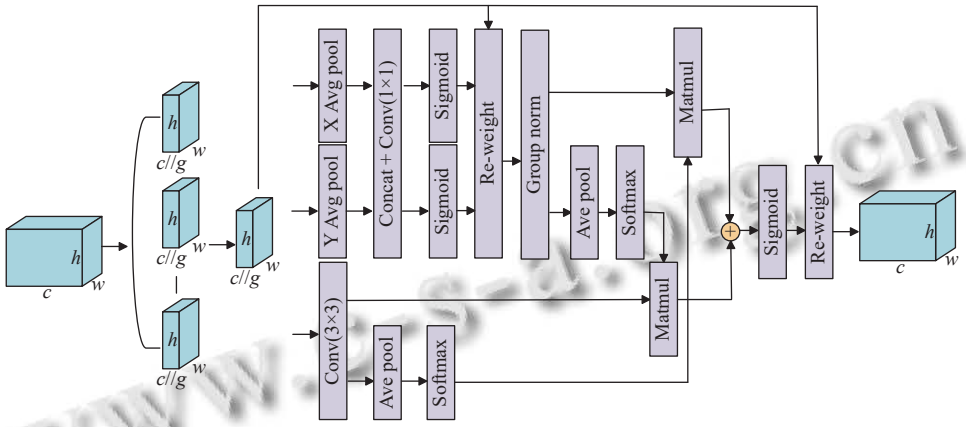

图 2 EMA 模块

EMA 因其独特的跨空间交互策略以及建立通道间关系的方式高效地捕获了不同方向上的长程依赖, 以及不同尺度下的特征信息. 因此, 本文引入 EMA 来实现对不同尺度下的特征捕捉以及载体图像和数字水印的特征融合.

1.2.2 金字塔多尺度跨时空注意力模块 (PEMA)

Zhao 等人[15]提出了金字塔池化模块 (pyramid pooling module), 该模块聚合了不同区域的上下文信息, 从而获取全局信息. 受其启发, 本文提出了金字塔多尺度跨时空注意力模块 PEMA (pyramid EMA) 来捕获不同分辨率下的特征并将其聚合, 以此来弥补单一分辨率处理情况下所带来的特征学习不充分的问题. PEMA 由 4 个分支组成, 每个分支分别将输入特征向量以不同的尺度平均划分, 通过对每个子区域实施 EMA, 得到不同分辨率与不同尺度下的局部注意力表示. 之后, 对 4 个分支的输出进行拼接融合, 以此来聚合不同分辨率与不同尺度下的全局注意力. PEMA 架构如图 3 所示.

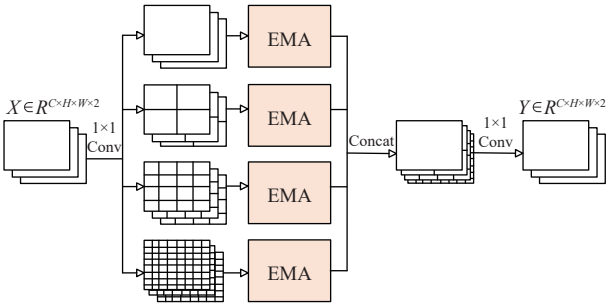

图 3 PEMA 架构

在输入到 PEMA 之前, 我们将之前得到的载体图像与数字水印的特征张量在 W 维度上拼接成一个特征张量$X \in R^{C\times H\times W\times 2}$, 一起送入到 PEMA 中处理. 特征张量进入 PEMA 之后, 首先被一个核为 1 的卷积处理, 之后分别送入 4 个分支中. 每个分支将特征张量平均分为$s\times s$个子区域$R_{s,i,j} \in R^{C\times \frac{H}{s}\times \frac{W}{s}\times 2}$, 其中$s \in \{1,2,4,8\}$, $1 \leqslant i,j \leqslant s$, 以得到不同分辨率下的特征. 之后在这 4 个分支中分别对每个子区域实施 EMA, 从而捕获不同尺度下的特征信息. 将这些子区域处理后得到的特征张量拼接起来便得到了该分支上的特征张量$Y_S \in R^{C\times H\times W\times 2}$. 将 4 个分支得到的输出在通道维度上拼接, 最后用一个核为 1 的卷积来聚合不同分辨率下不同尺度的时空特征, 得到最终的特征张量$Y \in R^{C\times H\times W\times 2}$.

1.3 离散小波变换

傅里叶变换可以分析信号的频谱, 但对于频率随着时间变化的非平稳信号, 其有着自身的局限性. 因其基函数为无限长的三角函数基, 故其存在吉布斯效应, 即用傅里叶变换去逼近存在突变的信号, 会在频率剧烈变化处产生突起. 因此傅里叶变换不能有效地表示突变信号. 为解决这个问题 STFT (short-time Fourier transform) 应运而生, 其通过加窗的操作, 把整个时域过程分解为无数个等长的小过程, 在每个小过程内信号可以近似看为平稳变化. 但其仍然存在一个问题, 即 STFT 的窗口是固定的, 窗太窄, 窗内的信号太短, 会导致频率分析不够精准, 频率分辨率差, 窗太宽, 时域上又不够精细, 时间分辨率低, 因此其无法满足非稳态信号变化的频率的需求.

自然界中的大量信号几乎都是非平稳的, 因此 DWT (discrete wavelets transform) 有着重要的意义. 小波变化是用特定的正交级数表示函数的一种方法, 其将无限长的三角函数基换成了有限长的会衰减的小波基, 以此来获得频域和时域上的信息, 常用的有小波基函数、Morlet 小波、Haar 小波等.

对于数字图像来说, 其由平滑部分和边缘构成, 其中边缘中包含着图像的大部分信息. 从频域上分析, 图像的低频部分有着其平滑部分的信息, 人眼对其不太敏感, 而高频部分有着其边缘部分的信息, 人眼感知上对其更为敏感. 故如果只对图像的高频部分进行修改, 人眼对其的感知并不敏感, 且图像的编码信息只会收到微小的损失. JPEG 压缩也正是基于此原理, 通过量化来压缩高频部分的信息, 从而达到保持低频分量、抑制高频分量的效果. 因此, 通过 DWT 对图像进行分解, 将数字水印嵌入到其低频分量上可以有效地抵御 JPEG 压缩攻击, 同时提高其在人眼感知上的不可见性.

在本文提出的网络中, 载体图像I_{co}首先经过 DWT 被分解为 4 个小波子带: LL, LH, HL, HH[16]. 其中, LL 包含着载体图像的低频信息, 选择其作为数字图像水印嵌入的载体. 需要注意的是, 在经过 DWT 处理后, 形状为$C\times H\times W$的载体图像会变为$C\times \frac{H}{2}\times \frac{W}{2}$, H 和 W 维度的减小也为后续的操作起到了降低运算开销的作用.

1.4 损失函数

本文提出的网络有两个主要的目标, 即编码图像在人眼感知上的不可见性, 以及解码器复原的数字水印的准确性. 模型采用的损失函数也是基于这两个主要目标来设计.

对于编码图像在人眼感知上的不可见性, 需要保证载体图像和编码图像的相似性. 因此, 对于编码器来说, 本文利用均方误差 (mean square error, MSE) 来计算载体图像和编码图像之间插值平方和的均值, 其公式如式 (2) 所示:

$$L_E = \frac{1}{N}\sum_{i=1}^{N}(I_{\text{en}} - I_{\text{co}})^2 \tag{2}$$

对于对抗判别器来说, 其主要目的是识别和区分载体图像I_{co}与嵌入水印后的图像I_{en}之间的细微差异以对抗编码器, 而编码器应该在训练中学习欺骗对抗判别器做出错误的判断, 故用二元交叉熵 (binary cross entropy, BCE) 来计算对抗判别器的结果是否为错误的, 其公式如式 (3) 所示:

$$L_A = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(D(I_{\text{en}})_i) + (1-y_i)\log(1-D(I_{\text{en}})_i)] \tag{3}$$

其中, $D(I_{\text{en}})$表示编码图像输入到对抗判别器后得到的

预测标签, y表示真实标签.

对于解码器来说, 其在训练过程中需要保证复原的数字水印的准确性, 故利用 MSE 计算数字水印和复原水印的差距, 其公式如式 (4) 所示:

$$L_D = \frac{1}{N}\sum_{i=1}^{N}(M' - M)^2 \tag{4}$$

除此之外, 为了保证编码图像的高隐蔽性, 经过 DWT 处理后的载体图像的 LL 子带应该与经过 DWT 处理后的编码图像的 LL 子带有着高相似性, 故用 MSE 来计算两者之间的差距, DLL (DWT LL sub-band loss) 的定义如式 (5) 所示:

$$L_{\mathrm{DLL}} = \frac{1}{N}\sum_{i=1}^{N}(F(I_{\mathrm{en}})_{\mathrm{LL}} - F(I_{\mathrm{co}})_{\mathrm{LL}})^2 \tag{5}$$

其中, $F(x)_{\mathrm{LL}}$代表进行 DWT 操作后取其 LL 子带.

最后, 本文提出的网络的总损失函数L_{total}的定义如式 (6) 所示:

$$L_{\mathrm{total}} = \lambda_E L_E + \lambda_A L_A + \lambda_D L_D + \lambda_{\mathrm{DLL}} L_{\mathrm{DLL}} \tag{6}$$

图 4 展示了本文提出的网络分别在岩心图像数据集和 COCO 公开数据集上训练的损失收敛曲线, 其中 loss 值代表每个 epoch 的平均批处理损失.

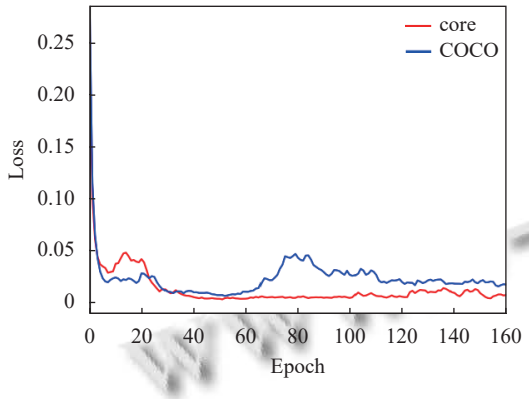

图 4 损失收敛曲线

2 实验结果与分析

2.1 评价指标

数字图像水印有两个主要任务: 鲁棒性用误码率 (bit error rate, *BER*) 来衡量; 隐蔽性用峰值信噪比 (peak signal-to-noise ratio, *PSNR*)[17]和结构相似性 (structural similarity, *SSIM*)[18]来衡量. 以上指标的计算公式分别为:

$$BER = \frac{N_e}{N} \tag{7}$$

其中, N表示发送的总比特数, N_e表示发生错误的比特数.

$$PSNR = 10\log_{10}\left(\frac{Max^2}{\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I(i,j) - K(i,j)]^2}\right) \tag{8}$$

其中, M 和 N 分别是图像的宽度和高度, $I(i,j)$ 是原始图像在位置(i,j)处的像素值, $K(i,j)$ 是编码图像在位置(i,j)处的像素值.

$$SSIM = \frac{(2\mu_x\mu_x + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + 2)} \tag{9}$$

其中, x 和 y 分别表示两幅图像的窗口, μ_x和μ_y分别表示窗口 x 和 y 的平均值, σ_x^2和σ_y^2分别表示窗口 x 和 y 的方差, σ_{xy}表示窗口 x 和 y 的协方差, C_1和C_2是用于稳定公式的小常数, 用以避免除零问题.

2.2 数据集介绍

本文所采用的数据集包括 COCO 数据集[19]和岩心图像数据集. 其中岩心图像数据集是由 PL8KCL-30KF 相机采集得到的, 包括不同类型的岩心图像总共 10 000 张. PL8KCL-30KF 相机专门设计用于工业自动化检测领域的高速线扫描应用, 由 I-TEK 自主研发和生产. 相机内部集成了国际领先的高速线阵图像传感器 (分辨率 8320×16), 大规模现场可编程逻辑门阵列并行处理器, 通过高速工业 Camera Link 总线对外输出数据, 同时具备了高速、高灵敏度和低噪声的特点. 而对于 COCO 数据集, 本文同样从中随机选取了 10 000 张图像用于实验. 本文将其中的 70% 用于训练, 20% 用于验证, 10% 用于测试.

为了保证岩心图像的原始性与权威性, 需要保存其原始图片, 故需将水印添加于原图之上. 由于实际中的岩心图像分辨率较大, 将其直接输入网络中会导致显存爆炸, 所以本文先将采集得到的原始岩心图像随机裁剪成 128×128 大小. 在应用时, 将岩心图像分割为 128×128 大小后添加水印, 拼接后形成编码图像, 其解码效果不会遭到影响.

2.3 实验参数

实验环境为 Ubuntu 18.04.6 LTS 系统, 处理器为

Intel(R) Xeon(R) CPU E5-2686 v4@2.30 GHz, GPU 为 NVDIA GeForce RTX 3090, 软件环境为 CUDA 11.1, 深度学习框架为 Torch 1.8.1.

对于损失函数的权重系数来说, 本文选择$\lambda_E=1$, $\lambda_A=0.0001$, $\lambda_D=10$, $\lambda_{\mathrm{DLL}}=0.1$. 在所有实验的测试过程中, 均采用质量系数$Q=50$的真实 JPEG 压缩. 对于梯度下降方面, 本文使用 Adam 优化器, 并设置初始学习率为 0.001. 批量大小为 16, 模型训练的 epoch 为 135.

2.4 对比实验

为探究本文所提算法的性能, 本文选取 HiDDeN、pseudo-differentiable deep learning approach[20](以下简称 Pseudo)、MBRS 和 CIN[21]作为基线模型进行对比实验, 每组实验均在相同的参数和设置下进行. 需特别注意的是, CIN 提供了其训练好的预训练模型, 其规定了载体图像大小为128×128, 水印长度为 30, 而本文所提供算法中数字水印信息 M 和载体图像I_{co}需要满足式 (1), 故在与其进行对比实验时将载体图像大小设置为192×192, 水印长度设置为 36, 均大于 CIN 的设置, 即在该设置下本文算法完成任务的难度要大于 CIN.

在岩心数据集中进行的实验结果及其参数如表 2 所示.

表 2 不同方法在岩心图像测试集上的指标

实验	模型	图像大小	水印长度	*BER*↓ (%)	*PSNR*↑ (dB)	*SSIM*↑
实验1	HiDDeN	128×128	64	16.79	33.1007	0.8644
	Pseudo	128×128	64	31.72	35.1408	0.8869
	MBRS	128×128	64	0.0875	39.6930	0.9578
	本文	128×128	64	**0.0140**	**42.0473**	**0.9774**
实验1	CIN	128×128	30	11.99	42.7712	—
	本文	192×192	36	**0.2388**	**45.1862**	**0.9878**

从表 2 中可以看出, 对于载体图像大小为128×128, 水印长度为 64 的情况, 本文所提出的算法在 *BER*、*PSNR* 以及 *SSIM* 指标上均优于现有主流模型. 而在与 CIN 进行对比时, 本文将载体图像大小设置为192×192, 水印长度设置为 36, 在这两方面更高的参数也使得本文所提供算法完成任务的难度更高, 但其 *BER*、*PSNR* 均高于 CIN, 进一步展示了本文算法的性能的优越. 由于 CIN 并未提供 *SSIM* 的指标, 故在此不进行该项指标的比较.

为验证本文所提出算法的泛化性, 本文在 COCO 公开数据集中进行了实验, 其结果及参数如表 3 所示. 可以看出, 除了在所针对的岩心数据集中表现优越, 本文所提供的算法在公开数据集 COCO 中同样有着不逊的表现. 在两种设置参数的环境下, 本文所提供算法的 *BER*、*PSNR* 以及 *SSIM* 指标均优于现有主流算法, 进一步验证了该算法良好的泛化性.

表 3 不同方法在 COCO 测试集上的指标

实验	模型	图像大小	水印长度	*BER*↓ (%)	*PSNR*↑ (dB)	*SSIM*↑
实验1	HiDDeN	128×128	64	22.31	31.6595	0.8810
	Pseudo	128×128	64	27.41	32.0007	0.8691
	MBRS	128×128	64	0.6906	37.9570	0.9601
	本文	128×128	64	**0.6515**	**40.3606**	**0.9620**
实验2	CIN	128×128	30	6.11	41.0343	—
	本文	192×192	36	**0.3805**	**42.8446**	**0.9826**

各算法在岩心图像测试集和 COCO 测试集上的效果对比分别如图 5、图 6 所示. 图中从上到下分别为载体图像、编码图像、编码图像和载体图像的差值、差值的归一化结果.

2.5 消融实验

为了验证引入 PEMA 模块、DWT 变换后的L_{DLL}损失函数以及联合损失函数后模型对于岩心图像嵌入鲁棒水印的有效性, 本文添加了消融实验, 载体图像大小为128×128, 水印长度为 64, 每组实验均在相同的参数以及设置下进行, 其结果如表 4 所示.

从表 4 中可以看出, PEMA 的加入, 使得的误码率降低了 0.068 8%, *PSNR* 提升了 1.482 4 dB, *SSIM* 提升了 0.014 1. L_{DLL}的加入, 使得误码率降低了 0.046 9%, *PSNR* 提升了 1.858 4 dB, *SSIM* 提升了 0.016. PEMA 和L_{DLL}同时引入使得误码率降低了 0.073 5%, *PSNR* 提升了 2.354 2 dB, *SSIM* 提升了 0.019 5.

此外, 本文还给出了损失函数中各部分的消融实验. 可以看出, 无论缺少了哪一部分, 都会导致最终的效果显著下降. 其中, L_D和L_A的缺失更是导致了误码率分别上升为 49.581 2% 和 43.184 6%, 概率分布已接近随机判断, 不可使用. 相关消融实验证明了联合损失函数的有效性.

从结果中我们可以发现, PEMA 因其捕获不同分辨率下多尺度特征信息的能力以及独特的跨空间交互策略使得网络能够更为高效地提取出特征张量的注意力描述, 为数字水印的嵌入提供了强大支撑. 而 DWT 操作以及L_{DLL}损失函数的引入使得数字水印能够更好地嵌入载体图像的低频区域中, 由于人眼对于低频信息较不敏感, 数字水印的隐蔽性也得到了保障. 二者的结合使得网络的效果得到了显著提升.

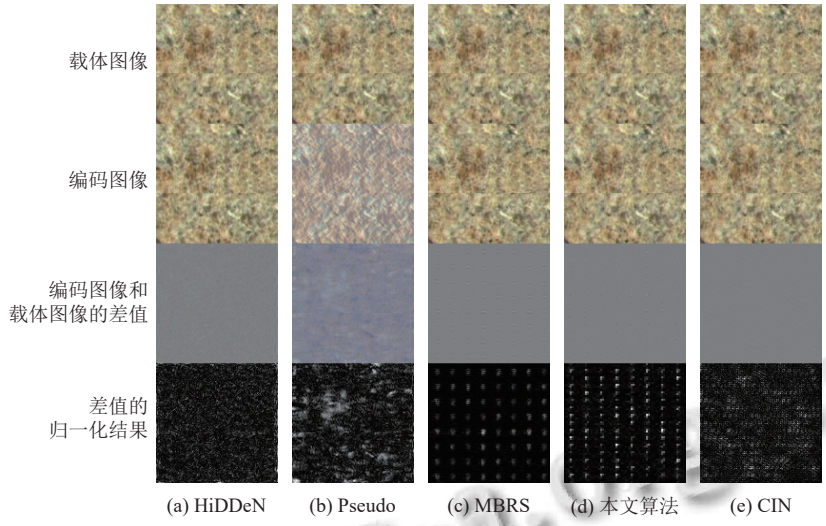

图 5 各算法在岩心图像测试集上的效果

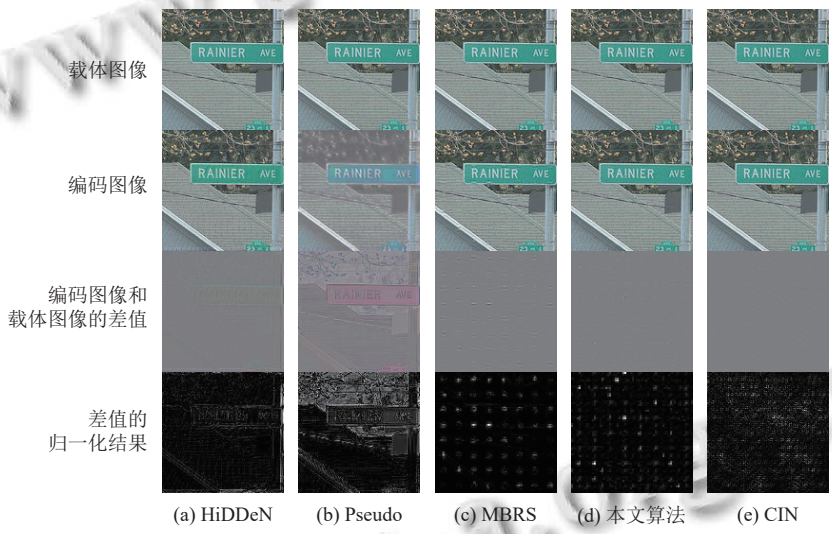

图 6 各算法在 COCO 测试集上的效果

表 4 消融实验

PEMA	L_{DLL}	L_E	L_D	L_A	*BER*↓ (%)	*PSNR*↑ (dB)	*SSIM*↑
—	—	√	√	√	0.0875	39.6931	0.9579
√	—	√	√	√	0.0187	41.1755	0.9720
—	√	√	√	√	0.0406	41.5515	0.9739
√	√	—	√	√	9.6641	39.2961	0.9471
√	√	√	—	√	49.5812	42.0388	0.9733
√	√	√	√	—	43.1846	41.0273	0.9641
√	√	√	√	√	**0.0140**	**42.0473**	**0.9774**

3 结语

本文提出了一个端到端岩心图像鲁棒水印算法来实现对于岩心图像的鲁棒性水印嵌入, 以此来抵抗针对 JPEG 压缩所带来的攻击. 本文引入了 PEMA 模块, 以其强大的捕获能力获取不同分辨率下多尺度的特征信息, 配合上其独特的跨空间交互策略, 为数字水印的嵌入提供了强大支撑, 大大降低了误码率. 而 DWT 操作以及 DLL 损失函数的引入利用人眼对低频信息不敏感的特性, 引导数字水印嵌入到载体图像的低频分量中, 保证了数字水印在人眼感知上的不可见性, 同时

也降低了运算开销. 本文进行的消融实验验证了每个模块的有效性, 同时, 和几个基线模型进行的对比实验也验证了本文提出算法的优越性能, 在公开数据集 COCO 上的实验也进一步验证了本文所提出算法的泛化性. 实验结果表明, 本文提出的算法相对于其他算法具有一定的优越性.

参考文献

1 Li MJ, Liu Y, Tian ZH, *et al*. Privacy protection method based on multidimensional feature fusion under 6G networks. IEEE Transactions on Network Science and Engineering, 2023, 10(3): 1462–1471. [doi: 10.1109/TNSE.2022.3186393]

2 Wang ZH, Byrnes O, Wang H, *et al*. Data hiding with deep learning: A survey unifying digital watermarking and steganography. IEEE Transactions on Computational Social Systems, 2023, 10(6): 2985–2999. [doi: 10.1109/TCSS.2023.3268950]

3 吕敏, 卿粼波, 滕奇志, 等. 基于 DCT 和 SVD 的岩心图像盲水印算法. 计算机与数字工程, 2016, 44(3): 515–520. [doi: 10.3969/j.issn.1672-9722.2016.03.029]

4 Zhang CN, Lin CG, Benz P, *et al*. A brief survey on deep learning based data hiding. arXiv:2103.01607, 2021.

5 Wallace GK. The JPEG still picture compression standard. Communications of the ACM, 1991, 34(4): 30–44. [doi: 10.1145/103085.103089]

6 Zhu JR, Kaplan R, Johnson J, *et al*. HiDDeN: Hiding data with deep networks. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 657–672.

7 Ahmadi M, Norouzi A, Karimi N, *et al*. ReDMark: Framework for residual diffusion watermarking based on deep networks. Expert Systems with Applications, 2020, 146: 113157. [doi: 10.1016/j.eswa.2019.113157]

8 Tancik M, Mildenhall B, Ng R. Stegastamp: Invisible hyperlinks in physical photographs. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 2117–2126.

9 Fang H, Jia ZY, Ma ZH, *et al*. PIMoG: An effective screen-shooting noise-layer simulation for deep-learning-based watermarking network. Proceedings of the 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022. 2267–2275.

10 Liu Y, Guo MX, Zhang J, *et al*. A novel two-stage separable deep learning framework for practical blind watermarking. Proceedings of the 27th ACM International Conference on Multimedia. Nice: ACM, 2019. 1509–1517.

11 Jia ZY, Fang H, Zhang WM. MBRS: Enhancing robustness of DNN-based watermarking by mini-batch of real and simulated jpeg compression. Proceedings of the 29th ACM International Conference on Multimedia. ACM, 2021. 41–49.

12 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.

13 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13713–13722.

14 Ouyang DL, He S, Zhang GZ, *et al*. Efficient multi-scale attention module with cross-spatial learning. Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes Island: IEEE, 2023. 1–5.

15 Zhao HS, Shi JP, Qi XJ, *et al*. Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2881–2890.

16 Singh K, Ranade S K, Singh C. Comparative performance analysis of various wavelet and nonlocal means based approaches for image denoising. Optik, 2017, 131: 423–437. [doi: 10.1016/j.ijleo.2016.11.055]

17 Almohammad A, Ghinea G. Stego image quality and the reliability of PSNR. Proceedings of the 2nd International Conference on Image Processing Theory, Tools and Applications. Paris: IEEE, 2010. 215–220.

18 Wang Z, Bovik AC, Sheikh HR, *et al*. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. [doi: 10.1109/TIP.2003.819861]

19 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. Proceedings of the 15th European Conference on Computer Vision. Zurich: Springer, 2014. 740–755.

20 Zhang CN, Karjauv A, Benz P, *et al*. Towards robust data hiding against (JPEG) compression: A pseudo-differentiable deep learning approach. arXiv:2101.00973, 2020.

21 Ma R, Guo MX, Hou Y, *et al*. Towards blind watermarking: Combining invertible and non-invertible mechanisms. Proceedings of the 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022. 1532–1542.

(校对责编: 张重毅)