陈帅龙, 丁海洋, 陈子超. 基于 ZUC-CNN 卷积神经网络的双层图像信息隐藏算法[J]. 智能计算机与应用,2026,16(3):49-57. DOI:10.20169/j.issn.2095-2163.25102902

基于 ZUC-CNN 卷积神经网络的双层图像信息隐藏算法

陈帅龙, 丁海洋, 陈子超
(北京印刷学院 信息工程学院, 北京 102600)

摘 要: 针对目前深度学习算法的端对端的编码器-解码器结构在平衡其机密性与隐写效果之间的问题,提出了一种基于 ZUC-CNN 卷积神经网络的双层图像信息隐藏算法。其双层隐写算法首先通过祖冲之流密码使用初始的密钥进行密码学级的加密,对秘密信息的原始统计分布和空间相关性进行置乱,然后通过基于 U-Net 的 HidingNet 网络实现信息隐藏以及基于 CNN 和 RevealNet 网络实现信息恢复,最后通过祖冲之流密码结合初始密钥进行解密。实验结果表明,该算法实现了嵌入容量为 24 bpp 的大容量的隐写,并且在原始图像的恢复和秘密信息的提取方面的 PSNR 分别达到了 40.10 dB 和 37.12 dB,保证了嵌入图像的最小失真。

关键词: 图像信息隐藏; 双层隐写模型; U-Net ; ZUC; CNN
中图分类号: TP391.41 **文献标志码**: A **文章编号**: 2095-2163(2026)03-0049-09

Two-layer image information hiding algorithm based on ZUC-CNN convolutional neural network

CHEN Shuailong, DING Haiyang, CHEN Zichao
(School of Information Engineering, Beijing Institute of Graphic Communication, Beijing 102600, China)

Abstract: To address the challenge of balancing security and imperceptibility in contemporary end-to-end deep learning-based steganography, this paper proposes a two-layer image information hiding algorithm based on ZUC-CNN convolutional neural network. Firstly, the proposed two-layer steganographic algorithm employs the ZUC stream cipher with an initial key to perform a cryptographic-level encryption of the secret information. This initial step effectively disrupts the original statistical distribution and spatial correlation of the secret data. Subsequently, a U-Net and CNN-based architecture, comprising a HidingNet and a RevealNet, is utilized for information embedding and extraction. The final recovery of the secret information is achieved by decrypting the extracted data using the ZUC cipher with the same initial key. Experimental results demonstrate that the proposed algorithm achieves a high embedding capacity of 24 bits per pixel (bpp). Furthermore, it ensures minimal distortion in the container image and high fidelity in the recovered secret information, with Peak Signal-to-Noise Ratio (PSNR) values reaching 40.10 dB and 37.12 dB for the reconstructed cover image and the extracted secret image, respectively.

Key words: image steganography; two-layer steganographic model; U-Net; ZUC; CNN

0 引 言

信息加密技术和信息隐藏技术作为两大核心防护手段,被广泛应用于维护网络空间的信息安全。加密技术,作为传统而经典的防御策略,通过复杂的算法机制,将明文信息转化为看似杂乱无章的密文,有效阻挡了未经授权的访问企图。这些算法,如高级加密标准(AES)、RSA 等,通过数学难题确保信息即便在传输过程中被截获,也无法轻易解读,从而保证了信息的秘密性。

信息隐藏技术,作为信息安全领域的另一重要分支,采取了一种更为隐蔽的策略。该技术不直接加密信息,而是将敏感数据巧妙地嵌入到看似无关的载体中,如图片、音频或视频文件,这一过程也被

基金项目: 国家自然科学基金(62472040); 北京市教委科研计划(KM202110015004); 北京市高等教育学会项目(MS2023204); 北京市数字教育研究课题(BDEC2023619095); 2026 年北京印刷学院本科教学改革创新项目(工程认证背景下 AI 赋能信息安全专业人才培养模式的创新与实践探索研究)。
作者简介: 陈帅龙(2001—),男,硕士研究生,主要研究方向:信息隐藏,数字水印; 陈子超(2002—), 男 ,硕士研究生,主要研究方向:信息隐藏。
通信作者: 丁海洋(1979—),男,博士,副教授,主要研究方向:数字水印,信息隐藏,密码技术应用。Email:dinghaiyang@bigc.edu.cn。
收稿日期: 2025-10-29

称为隐写术[1]。在现代隐写术中,目标是秘密地传达数字信息。隐写过程将隐藏的消息放置在称为载体的传输介质中。承运人可能是公开可见的。为了增加安全性,还可以对隐藏消息进行加密,从而增加感知的随机性并降低内容发现的可能性,即使检测到消息的存在也是如此[2]。

随着人工智能和机器学习的发展,信息隐藏技术也在不断进化,变得更加智能化和动态化,取得了比传统方法更好的性能。低容量信息隐藏的一个重要应用是知识产权保护,如数字水印[3-6]。Hayes 等学者[7]首次将 GAN 应用于信息隐藏任务,表明对抗训练方案可以有效提高隐藏安全性。同样,Shi 等学者[8]提出了 SSGAN,通过使用 WGAN[9]作为生成器、GNCNN[10]作为鉴别器的架构来提高隐写图像的质量。但是只能隐藏少量数据。对于数字水印来说,鲁棒性是最重要的因素,要求在不同类型的失真下都能准确地恢复秘密信息。然而,在鲁棒性和安全性之间存在权衡。虽然上述方法对失真具有鲁棒性,但通常安全性较低,即隐藏的信息很容易被第三方检测到。这一缺点使得上述方法不适合秘密通信。

与低容量信息隐藏工作相比,图像隐藏需要更高的容量。Baluja[11]首先提出使用深度神经网络将整个彩色图像隐藏在另一个图像中。为了实现这一目标,首先建立预备网络提取秘密图像的有用特征,然后利用隐藏网络将秘密图像的特征融合到封面图像中,但这增加了整个系统的重建误差,并增加了载体图像被隐写分析检测的可能性,然而这也为隐藏算法的探索开辟了一条新的途径。

为解决上述问题,本文提出了一种基于 ZUC-CNN 卷积神经网络的双层图像信息隐藏算法。通过祖冲之流密码[12]的秘密图像预处理模块,使用初始的密钥进行密码学级的加密,对秘密信息的原始统计分布和空间相关性进行置乱,然后通过图像隐写网络实现大容量的信息隐藏,最后通过图像提取模块以祖冲之流密码解密模块,使用初始的密钥实现对秘密图像的恢复。实验结果表明,该算法实现了嵌入容量为 24 bpp 的大容量的隐写结果,并且在原始图像的恢复和秘密信息的提取方面的 PSNR 分别达到了 40.10 dB 和 37.12 dB,保证了嵌入图像的最小失真。

1 相关工作

1.1 对抗神经网络隐写

为了摆脱对人工设计特征的依赖,文献[7]提出了一个开创性的三方对抗学习框架,用于自动生成隐写算法。该框架通过编码器(Alice)、解码器(Bob)和分析器(Eve)之间的博弈,成功地学习到了能够欺骗神经网络分析器的隐写策略。然而,该方法在训练过程中需要对 3 个网络的损失函数进行复杂的权重平衡,且其安全性主要在对抗一个同等架构的分析器时得到验证。当面对更强大或结构未知的外部隐写分析器时,其鲁棒性仍有待进一步探究。此外,该框架生成的隐写修改主要集中在图像的低频区域,这可能使其容易受到某些特定分析技术的攻击。

1.2 CNN 卷积神经网络隐写

深度学习的引入为图像隐写学开辟了新的范式。文献[11]中提出的深度隐写是该领域的开创性工作之一。该方法设计了一个包含预处理、隐藏和揭示网络的集成系统,通过最小化载体图像和秘密图像的重构误差($\|c-c'\|+\beta\|s-s'\|$)进行端到端训练。与依赖最低有效位(LSB)[13]的传统方法不同,文献[11]的模型学会了将秘密图像信息压缩并分布在载体图像的所有比特位中,从而在视觉上取得了较好的隐藏效果。延续这一思路,Rehman 等学者[14]也采用了基于 CNN 的编码器-解码器架构,证明了通过联合损失函数进行优化的有效性,并进一步提升了数据隐藏的性能。尽管上述基于深度学习的方法在隐藏容量方面远超传统算法,但其模型仍存在不足。文献[11]的研究明确指出,其模型在处理训练集之外的图像类型时表现不佳,例如在纯色背景中隐藏信息会产生肉眼可见的失真。此外,后续研究也观察到,在载体图像的低频(平滑)区域,信息嵌入操作容易引入不自然的噪声伪影,这降低了隐写的不可感知性。

1.3 U-Net 神经网络隐写

Van 等学者[15]提出了一种基于深度学习的图像隐写方法,其模型由一个用于嵌入的隐藏网络和一个用于提取的显示网络构成(具体见图 1)。该架构借鉴了 U-Net[16]的设计思想,其中每个卷积层后均采用了批归一化与 ReLU 激活函数。在网络输入端,该方法将载体图像与秘密图像在通道维度上进行拼接处理。为了优化模型性能,研究中联合使用了结构相似性(SSIM)和均方误差(MSE)两种损失函数进行约束。尽管该方法验证了编解码器框架在图像隐写中的有效性,但其报告的实验主要在低分辨率的图像上进行,并且输入端的直接拼接策略可能在抵抗高级隐写分析攻击时存在一定的安全脆弱性。

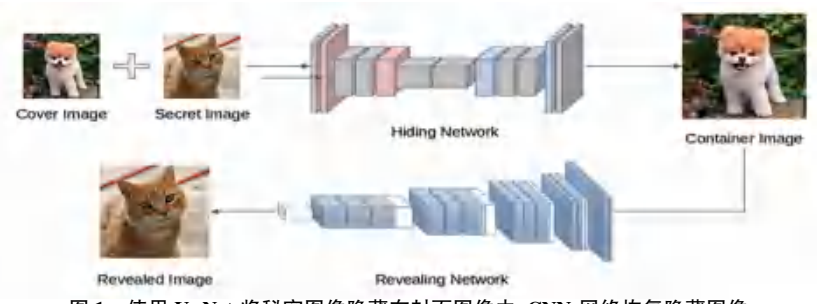

图 1　使用 U-Net 将秘密图像隐藏在封面图像中，CNN 网络恢复隐藏图像

Fig. 1　Hiding the secret image in the cover image with U-Net and restoring the hidden image with the revival CNN network

2　基于 ZUC-CNN 卷积神经网络的双层图像信息隐藏算法

本节详细阐述所提出基于 ZUC-CNN 卷积神经网络的双层隐写算法。该框架旨在通过在深度隐写模型中集成密码学的算法，提升秘密信息的安全性，同时保证图像隐写的不可感知性与鲁棒性。算法流程如图 2 所示，框架遵循“先加密后嵌入”的核心原则，其数据处理流程主要包含 3 个阶段：基于 ZUC 流密码的秘密图像预处理模块；基于 CNN 网络的隐写及提取模块；基于 ZUC 流密码的秘密图像解密模块。

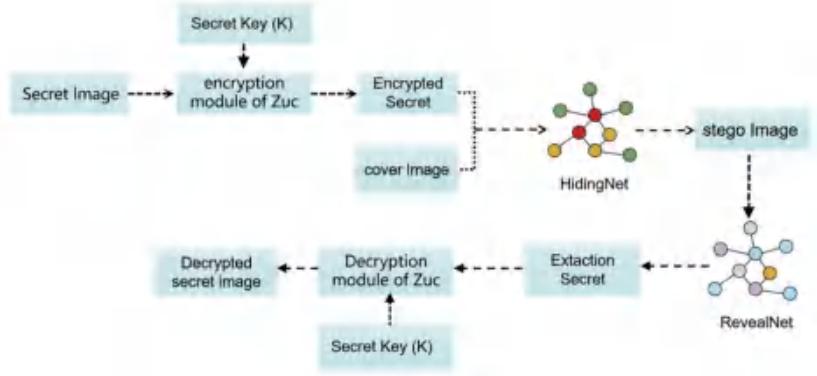

图 2　ZUC-CNN 卷积神经网络双层隐写整体流程图

Fig. 2　Flow chart of two-layer steganography in ZUC-CNN convolutional neural network

(1) 密码学预处理。原始秘密图像经归一化与量化后输入 ZUC 流密码加密模块；该模块在预共享密钥与初始向量的控制下生成伪随机密钥流，对图像按位进行一一对应的异或运算，输出加密秘密图像。

(2) 在嵌入阶段，系统将加密秘密图像与载体图像并行输入深度隐藏网络。该网络通过下采样和特征提取，学习在保证隐写图像与原始载体图像在视觉与统计域高度一致的约束下，将密文扰动以低可感知方式扩散嵌入。

(3) 信息提取端与嵌入结构互逆。接收端获取隐写图像，输入还原网络，经多重卷积输出对秘密图像的提取。

该“加密→嵌入→提取→解密”串联结构，一方面通过对秘密图像统计分布的前置随机化，显著削弱了传统隐写分析中依赖的跨域相关特征；另一方面，将攻击难度分解为“成功解嵌”与“成功解密”两个逻辑独立子问题，提升总体攻破复杂度。

2.1　算法主要步骤

2.1.1　隐藏网络 HidingNet 结构

隐藏网络部分，研究中选用了 U-Net 架构作为核心，HidingNet 是一个编码器-解码器结构。一个下采样的编码器，负责提取载体图像和秘密图像的多尺度特征，以及一个上采样的解码器，通过跳跃连接机制将高层次语义信息与低层次细节结合，实现对秘密信息的精确嵌入。这一过程中，秘密图像的数据被巧妙地与原始载体图像的特征融合，生成的载密图像在视觉上几乎无法与载体图像区分开来，但内含了加密版权等信息。

分析可知，编码器部分由多个卷积层和池化层组成，来逐步地降低图像的空间分辨率，并在降低的同时，增加通道数来捕捉更高层次的特征。

(1) 卷积层可学习的卷积核 K 是有于提取图像 x 的局部特征，对此可表示为：

$$(K * x)_{i,j} = \sum_{m} \sum_{n} K_{m,n} \cdot x_{i+m,j+n} \qquad (1)$$

其中，x 表示输入图像；$(K*x)_{i,j}$ 表示卷积后的特征图的一个元素；i 和 j 是特征图的索引。此外，在本文中的“$*$”都为卷积操作。

并且在除最后一层之外，其卷积层都后接一个 ReLU 激活函数，增强其网络的表达能力。

(2)在经过几次的卷积层后，为了捕捉更高层次的特征，降低图像的空间维度(高度以及宽度)，会先通过批量归一化来规范化卷积层的输出，提高训练稳定性，其公式为：

$$\mathrm{BN}(a)=\gamma\left(\frac{a-\mu}{\sigma}\right)+\beta \tag{2}$$

其中，μ 表示特征图 a 的均值，σ 表示标准差，γ 和 β 表示可学习的参数。然后通过其池化层进行池化操作，对此可以表示为：

$$P(a)=\max_{i,j}(a(i,j)) \tag{3}$$

使用最大池化，通过每个池化窗口中取得最大值来得到最后的池化后的特征图。

(3)通过跳跃连接将编码器中的特征图直接连接到解码器的对应层，有助于在图像重建过程中保留更多的细节信息。这里用到的公式为：

$$S(x_e,x_d)=x_d+F(x_e) \tag{4}$$

其中，x_e 表示编码器中的特征图；x_d 表示解码器中的特征图；F 表示一个函数，是一个 1×1 的卷积操作。

(4)在解码器(上采样)部分之中，由上采样(转置卷积)层组成，可在逐步恢复图像的空间维度的同时，减少通道数。转置卷积的基本思想是通过使用转置的卷积核在特征图上进行操作，以增加其空间尺寸。转置卷积一个学习填充降采样过程中丢失信息的过程。转置卷积的数学表达式为：

$$(K^{\mathrm{T}}*F)_{i,j}=\sum_m\sum_n K^{\mathrm{T}}_{m,n}\cdot F_{i+m,j+n} \tag{5}$$

其中，K^{T} 表示转置后的卷积核；F 表示编码器中的特征图；$(K^{\mathrm{T}}*F)_{i,j}$ 表示上采样后的特征图的一个元素。通过这种方式，转置卷积层能够将特征图的尺寸从 $H\times W$ 增大到 $(H\times s)\times(W\times s)$，其中 s 是转置卷积的步长。经过式(5)的转置卷积操作之后，会接续一个批量归一化层，通过规范化转置卷积层的输出来稳定训练过程，并提高模型的泛化能力。批量归一化层会对每个特征通道的输出进行归一化，使其具有均值为 0、方差为 1 的分布，然后通过 2 个可学习的参数 γ 和 β 进行缩放和平移变换(具体见式(2))。

在上采样的过程中，网络逐渐重建了图像的空间分辨率，同时通过跳跃连接保留了编码器中的细节信息。这一过程对于生成视觉上与载体图像相似的隐写图像至关重要，因为在允许网络在嵌入秘密信息的同时，重建图像的局部特征和纹理。至此，解码器的最后几层将这些融合的特征图转换为最终的载密图像。

通过这种方式，HidingNet 能够在不显著改变载体图像外观的情况下，实现秘密信息的有效隐藏。上采样过程中确保在图像重建的同时，秘密信息被嵌入到载体图像中，生成的隐写图像既具有高质量的视觉效果，又能够通过特定的方法被提取出来。其隐藏网络如图 3 所示，其步骤包括：

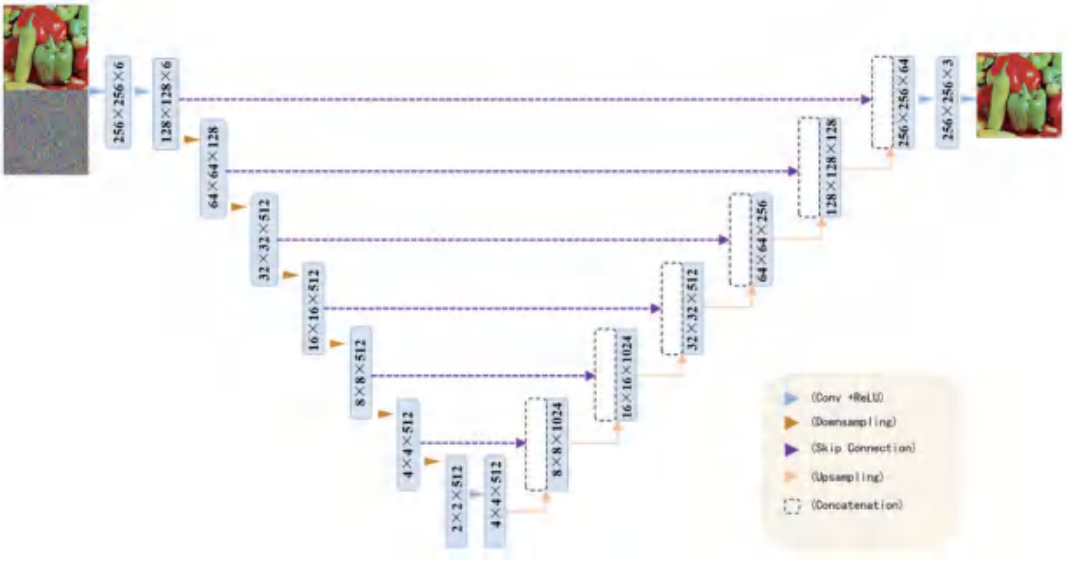

图 3 HidingNet 隐藏网络图

Fig. 3 HidingNet hidden network diagram

(1)将载体图像和经过祖冲之序列密码算法加密后的秘密图像输入到编码器块中进行逐步的下采样,直到将 256×256 的 6 通道特征图经过卷积层、激活函数以及 BN 层逐步的 2×2 的 512 通道的特征图。

(2)当完成下采样之后,将其下采样的特征图传输到解码器中进行上采样,通过反卷积、激活函数、BN 层以及与跳跃连接与下采样阶段的特征图进行拼接,最终将图像放大为一个 256×256 的图像作为输出。

2.1.2　水印提取网络 RevealNet

解码器网络部分,本文设计了一个全卷积网络(FCN),该网络由 6 层卷积层构成,有效捕获局部特征并逐步提炼信息。为了保证网络的非线性表达能力和训练稳定性,除最后一层外,所有卷积层之后均加入了 ReLU 激活函数以增加网络的表达力,并通过批量归一化减少内部协变量偏移。解码器网络的输出层使用 Sigmoid 激活函数,确保输出值在 0~1 之间,与图像像素值范围相符,从而能精确地恢复出隐藏的秘密图像。该网络的设计重点在于提取和处理隐写图像中的关键特征,以实现高质量的秘密信息恢复。

RevealNet 的架构基于卷积神经网络(CNN)。RevealNet 的输入是经过隐写术处理的图像。这些图像在视觉上看起来与原始图像相似,但实际上包含了隐藏的信息。RevealNet 的核心是多个卷积层,这些层负责从输入图像中提取特征。每个卷积层由一系列可训练的卷积核组成,这些卷积核在训练过程中学习到能够识别图像中特定模式的权重。其卷积操作如式(1)所示,通过多层卷积操作,RevealNet 能够捕捉到从低级到高级的特征表示。为了防止只能学习输入和输出之间的线性关系,在每个卷积层之后,RevealNet 应用一个非线性激活函数 ReLU,并同样使用批量归一化以及池化层来得到特征图,最后通过一个 1×1 的卷积层来生成最终的输出图像。提取网络流程如图 4 所示。图 4 中,所有卷积都是 3×3,批归一化层(BN),填充为 1,步长为 1。除了输出层是 Sigmoid 激活函数外,其他层都是 ReLU 激活函数。其具体步骤为:

(1)将用户提供的含有秘密图像的载密图像传入输入层之中。

(2)载密图像进行多轮的卷积层、BN 层以及激活层征处理,在本文中为 6 层卷积,最终分别输出提取的秘密图像以及提取秘密图像后的载体图像。

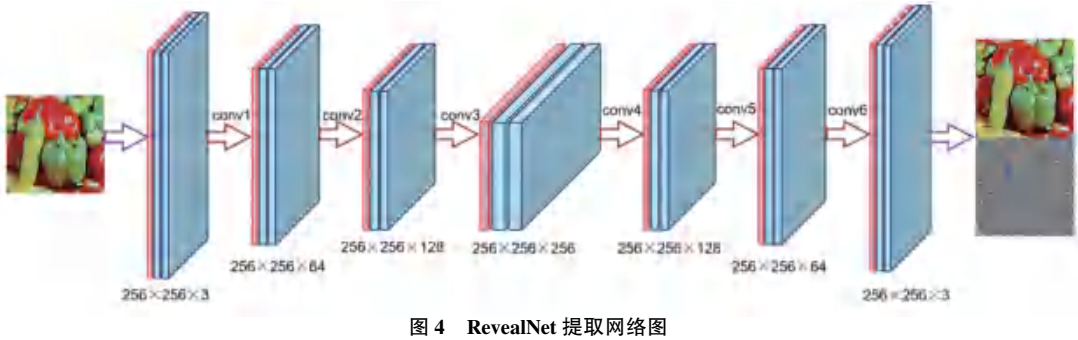

图 4　RevealNet 提取网络图

Fig. 4　RevealNet extraction network diagram

2.1.3　损失函数

为了衡量生成的载密图像 c' 与原始的载体图像 c 之间的差异,以及提取的秘密图像 s' 与原始秘密图像 s 之间的差异,研究中引入了损失函数。损失函数的目的是最小化这些差异,从而优化隐写网络和提取网络的学习过程。其中通过反向传播算法(Backpropagation)不断调整模型参数 $\Theta = \{w_i, b_i\}$,这里 w_i 和 b_i 分别表示网络中的权重和偏置。并使用均方误差(Mean Squared Error, MSE)来衡量网络重建图像 $F(Y;\Theta)$ 与真实图像 X_i 之间的差异。

对于单个训练样本,损失函数 $L(\Theta)$ 定义为重建图像 $F(Y;\Theta)$ 与真实图像 X_i 差的平方和的平均值。定义公式如下:

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{N} \| F(Y;\Theta) - X_i \|^2 \qquad (6)$$

其中,n 表示训练样本的数量。

对于整个数据集,损失函数 $L(c,c',s,s')$ 定义为原始载体图像和秘密图像的重建误差的加权和。定义公式为:

$$L(c,c',s,s') = \frac{1}{2}\| c - c' \|^2 + \alpha \| s - s' \|^2 \qquad (7)$$

其中,α 表示用于平衡 2 种误差的权重参数。

为了最小化损失函数,本文使用了 Adam 优化器,其结合了动量和 RMSProp 思想的优化算法,同时利用梯度的一阶和二阶矩来更新参数。对于参数更新遵循以下规则:

$$\Delta_{k+1} = 0.9 \times \Delta_k - \eta \Delta L \tag{8}$$

其中，Δ_{k+1} 表示在第 $k+1$ 步的参数更新；Δ_k 表示第 k 步的动量项；η 表示学习率；ΔL 表示损失函数 L 的梯度。动量项 $0.9 \times \Delta_k$ 有助于平滑更新过程，减少训练过程中的噪声。

最终，网络权重 W 根据更新值 Δ 进行调整：

$$W_{k+1} = W_k + \Delta_{k+1} \tag{9}$$

通过这种方式，网络在训练过程中可以不断调整参数，以最小化隐写图像和载密图像之间的差异，同时确保秘密信息能够被准确嵌入并从隐写图像中恢复出来。这个过程涉及到反向传播算法，可通过计算损失函数关于网络参数的梯度来指导参数更新。

总体而言，本研究提出的基于 ZUC-CNN 卷积神经网络的图像信息隐藏算法，不仅在技术上实现了对传统隐写术的重大突破，提高了图像水印的嵌入与提取效率，同时也展示了深度学习技术在信息安全领域、尤其是数字版权保护和隐私保护方面的广泛应用前景。通过持续优化网络结构和算法策略，以及与密码学的紧密结合，后续将进一步提升图像隐写的隐蔽性和鲁棒性，为未来安全通信和数字媒体版权管理开辟新途径。

3 实验

3.1 实验设置

本文从 ImageNet[17] 和 COCO[18] 数据集中各收集了 12 000 张用于训练的图像和 2400 张用于测试的图像作为训练集训练网络模型。网络的初始学习率设置为 0.001，超参数 α 设置为 0.75。采用 Adam 优化方法自动调整学习速率，使网络参数学习流畅。每批图像数设置为 32，网络训练 200 次迭代。本文的实验环境是在 Autodl 服务器上运行，GPU 是 NVIDIA GeForce 3090，实验环境是 Pytorch，开发环境为 pycharm，Python 版本为 Python 3.8 进行模拟实验，该模型的训练结果将从主观隐写结果和客观隐写能力两个方面验证所提方法的实用性。

3.2 评价指标

为了更好地对数字水印的嵌入效果进行评估，在本文中引用峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）、结构相似性（Structural Similarity Index，SSIM）来判定。

（1）峰值信噪比（PSNR）是衡量图像失真的重要指标。具体表示图像最大可能功率与噪声功率之比，通常用对数分贝单位表示。PSNR 可通过均方误差（MSE）计算，MSE 定义为 2 个 $m \times n$ 单色图像 I（无噪声的原始图像）和 K（I 的噪声近似）之间的差异，数学定义公式如下：

$$\mathrm{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{i,j} - K_{i,j}]^2 \tag{10}$$

PSNR 的定义为：

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) = 20 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}}}\right) \tag{11}$$

其中，MAX_I 表示图像点颜色的最大数值。

（2）结构相似性指数（SSIM）是一种关键的评估手段，用于判断 2 幅数字图像在内容上的相似度。该指标尤其适用于如下场景中：其中一幅图像遭受扭曲处理，而另一幅则保持原样无损。

给定 2 个信号 x 和 y，两者的结构相似性定义为：

$$\mathrm{SSIM}(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{12}$$

其中，$l(x,y)$ 比较 x 和 y 的亮度；$c(x,y)$ 比较 x 和 y 的对比度；$s(x,y)$ 比较 x 和 y 的结构，其计算公式分别为：

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{13}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{14}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{15}$$

3.3 实验结果分析

3.3.1 主观质量分析

图 5 是模型的输出图像效果。秘密图像通过隐藏网络将秘密信息嵌入载体图像后再通过重建载体图像生成载密图像，其中图 5（a）、图 5（b）为原始的载体图像，图 5（c）、图 5（d）为秘密图像，图 5（e）、图 5（f）为 ZUC 算法加密后的图像，其载密图像、即图 5（g）和图 5（h）看起来几乎与载体图像没有区别，视觉的差别很小，对于人的肉眼很难察觉并分辨出是否是有变化。并且提取出的秘密图像、即图 5（i）和图 5（j）与经过 ZUC 解密算法后恢复的秘密图像、即图 5（k）和图 5（l）在视觉上的变化也很小，对于图像的失真几乎感知不到。

3.3.2 客观质量分析

在 PSNR 和 SSIM 等指标外，本研究还通过分析像素直方图来从统计学层面评估算法的不可感知性。一个优秀的隐写算法应确保载密图像的像素值分布与原始载体图像的分布尽可能接近，以抵抗基于一阶统计特征的检测。

4 幅图像像素直方图如图 6 所示，具体展示了 2

组典型的实验结果及其对应的像素直方图。图 6 中，图 6(a)为原始载体图像，图 6(b)为原始的秘密图像，图 6(c)和图 6(d)分别对应图 6(a)和图 6(b)的像素直方图。图 6(e)为隐写图像，图 6(f)为提取后经过 ZUC 算法解密的秘密图像，而图 6(g)与图 6(h)则对应图 6(e)和图 6(f)的像素直方图。

图 5　模型的输出效果图

Fig. 5　Diagram of model output effect

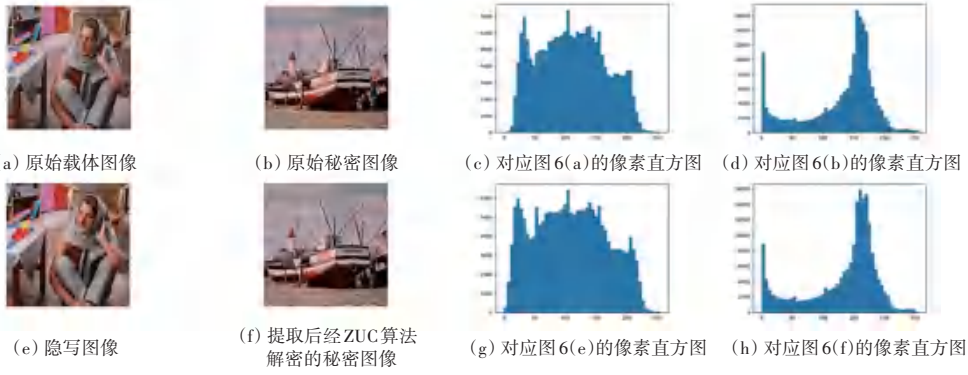

图 6　4 幅图像像素直方图

Fig. 6　Pixel histograms of four images

通过仔细比对可以发现，隐写图像(g)的直方图在整体轮廓、峰值位置和分布趋势上，与原始载体图像(a)的直方图保持了高度的一致性。同样地，直方图(h)也精确地复现了直方图(d)的统计分布特征。尽管嵌入了高达 24 bpp 的大容量数据，但信息嵌入过程对像素值的扰动非常微小。这一结果客观地证明了本算法所生成的载密图像能够很好地保持原始图像的一阶统计特性，从而具备较强的统计隐蔽性。其 PSNR 及 SSIM 见表 1。

为了验证本文中的模型具有可行性以及可靠性，对于原始的载体图像、原始载体图像-隐写图像以及原始秘密图像-解密后的秘密图像进行了 PSNR 与 SSIM 的分析。由表 1 结果可知，在原始图像的恢复和秘密信息的提取方面，PSNR 分别达到了 40.10 dB 和 37.33 dB，保证了嵌入图像的最小失真。

表 1　PSNR 及 SSIM 对比表

Table 1　Comparison table of PSNR and SSIM

比较的图片	PSNR/dB	SSIM
第一组原始载体图像-隐写图像	40.104 5	0.978 9
第一组原始秘密图像-解密后的秘密图像	35.231 0	0.961 7
第二组原始载体图像-隐写图像	39.239 4	0.980 1
第二组原始秘密图像-解密后的秘密图像	37.326 1	0.982 6

为了客观评估所提算法的性能，本研究将其与多种先进的深度学习隐写方法进行了比较，结果见表 2。这些方法涵盖了不同的技术路线，包括经典的编码器-解码器架构、生成对抗网络。

表 2 与其他算法 PSNR 对比表

Table 2 Comparison of PSNR with other algorithms

Method	Technique	Cover Image Size	Secret Image Size	PSNR
Rehman's method[19]	Encoder-decoder	32×32×3	32×32×1	32.90
Zhang's method[20]	GAN	256×256×3	256×256×1	33.92
Wang's method (ISGAN)[21]	GAN	256×256×3	256×256×1	34.01
Yang's method[22]	ISGAN	300×300×3	300×300×1	34.07
Subramanian's method[23]	Encoder-decoder	256×256×3	256×256×3	34.55
Proposed method	Encoder-decoder	256×256×3	256×256×3	37.33

(1)在处理相同尺寸(256×256)的载体图像时,本研究提出的方法在载密图像的峰值信噪比(PSNR)上达到了 37.33 dB,显著优于所有对比方法。

(2)与同样采用编码器-解码器架构的 Subramanian 等学者[23]的方法相比,本方法在嵌入容量完全相同的情况下,PSNR 提升了 1.06 dB。这表明本文所设计的基于 U-Net 的隐藏网络在最小化载体图像失真方面具有更强的能力。

(3)与基于 GAN 的方法、如 Wang 等学者[21]和 Yang 等学者[22]相比,本方法不仅在 PSNR 上分别高出 1.60 dB 和 1.54 dB,更重要的是,实现了从隐藏单通道灰度图到隐藏三通道彩色图的跨越,嵌入容量提升至原来的 3 倍。

这一结果有力地证明了本算法在实现大容量信息隐藏的同时,依然能保持强大的不可感知性。

4 结束语

本文主要研究了基于 ZUC-CNN 卷积神经网络的双层图像信息隐藏算法,并进行了综合分析和实验验证。实验结果表明,所提出的数字水印方案具有较高的隐藏容量以及较好的隐藏效果,在视觉上其失真的效果得到了改善。本文研究的数字水印嵌入算法,实现了以图藏图的方式,旨在解决数字图像在网络传播过程中可能面临的版权侵权和篡改问题,为保护知识产权提供有效的技术手段和方法。研究说明数字水印技术与深度学习结合的有效性和潜力,同时也发现数字水印技术在实际应用中仍然存在挑战和限制。未来可以进一步优化数字水印算法,提升其对秘密图像的隐藏容量以及优化隐藏效果。

参考文献

[1] RUSTAD S, ANDONO P N, SHIDIK G F. Digital image steganography survey and investigation(goal, assessment, method, development, and dataset)[J]. Signal Processing,2023,206:108908.

[2] JOHNSON N F, DURIC Z, JAJODIA A S. Information hiding: Steganography and watermarking - attacks and countermeasures [M]. Cham: Springer, 2001.

[3] ZHU Jiren, KAPLAN R, JOHNSON J, et al. Hidden: Hiding data with deep networks [C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 657-672.

[4] AHMADI M, NOROUZI A, KARIMI N, et al. ReDMark: Framework for residual diffusion watermarking based on deep networks[J]. Expert Systems with Applications, 2020,146:113157.

[5] TANCIK M, MILDENHALL B, NG R. Stegastamp: Invisible hyperlinks in physical photographs[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE, 2020: 2117-2126.

[6] LUO Xiyang, ZHAN Ruohan, CHANG Huiwen, et al. Distortion agnostic deep watermarking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE, 2020: 13548-13557.

[7] HAYES J, DANEZIS G. Generating steganographic images via adversarial training [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach,USA: NIPS Foundation, 2017, 30: 1951-1960.

[8] SHI Haichao, DONG Jing, WANG Wei, et al. SSGAN: Secure steganography based on generative adversarial networks [C]// Pacific Rim Conference on Multimedia. Cham: Springer, 2018, 10735: 534-544.

[9] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks [C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: PMLR, 2017,70: 214-223.

[10] QIAN Yinlong, DONG Jing, WANG Wei, et al. Learning and transfer representations for image steganalysis using convolutional neural network [C]//Proceedings of 2016 IEEE International Conference on Image Processing (ICIP). Piscataway,NJ:IEEE, 2016: 2752-2756.

[11] BALUJA S. Hiding images in plain sight: Deep steganography [J]. Advances in Neural Information Processing Systems, 2017, 30: 2066 - 2076.

[12] FRANCO G, MARGENSTERN M. ADNA computing inspired computational model[J]. Theoretical Computer Science,2008,404(1-2):99-96.

[13] MIELIKAINEN J. LSB matching revisited [J]. IEEE Signal Processing Letters, 2006, 13(5): 285-287.

[14] REHMAN A U, RAHIM R, NADEEM S. End-to-end trained CNN encoder-decoder networks for image steganography [C]// Proceedings of the European Conference on Computer Vision (ECCV) Workshops. Cham:Springer, 2019: 11132:723-729.

[15] VAN T P, DINH T H, THANH T M. Simultaneous convolutional neural network for highly efficient image steganography [C]// Proceedings of 2019 19th International Symposium on Communications and Information Technologies (ISCIT). Piscataway, NJ: IEEE, 2019: 410-415.

[16] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]// Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015, 9351: 234-241.

[17] RUSSAKOVSKY O, DENG Jia, SU Hao, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[18] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft Coco: Common objects in context [C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2014: 740-755.

[19] REHMAN A U, RAHIM R, NADEEM S. End-to-end trained CNN encoder-decoder networks for image steganography [C]// Proceedings of the European Conference on Computer Vision (ECCV) Workshops. Cham: Springer, 2019: 723-729.

[20] ZHANG K A, CUESTA-INFANTE A, XU Lei, et al. SteganoGAN: High capacity image steganography with GANs [J]. arXiv preprint arXiv, 1901.03892, 2019.

[21] WANG Zihan, GAO Neng, WANG Xin, et al. STNet: A style transformation network for deep image steganography [C]// Proceedings of International Conference on Neural Information Processing. Cham: Springer, 2019: 3-14.

[22] YANG Wenxin, LI Longjie, BAI Shenshen, et al. IS-GNN: Graph neural network enhanced by aggregating influential and structurally similar nodes [J]. Knowledge-Based Systems, 2024, 301: 112-282.

[23] SUBRAMANIAN N, CHEHEB I, ELHARROUSS O, et al. End-to-end image steganography using deep convolutional autoencoders [J]. IEEE Access, 2021, 9: 135585-135593.