Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2024, 13(4), 765-771
Published Online November 2024 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2024.134078

Hans汉斯

基于编码－解码网络图像信息隐藏算法

李雅静，丁海洋*，熊　涛

北京印刷学院信息工程学院，北京

收稿日期：2024年10月8日；录用日期：2024年11月7日；发布日期：2024年11月20日

摘　要

传统的图像隐写往往倾向于将隐藏信息安全地嵌入到封面图像中，而几乎忽略了有效负载容量。为解决传统隐写容量低的问题，本文采用深度学习与图像信息隐藏相结合的方法。实验结果表明，在嵌入容量上，所提算法达到了24 bpp，是目前容量最大的图像隐写算法之一。在此大容量嵌入的前提下，所提算法生成的载密图像和提取的秘密图像，无论在主观视觉质量还是客观视觉指标峰值信噪比(PSNR)上都高于其他同类算法，说明了设计的端到端隐写网络的整体优越性。

关键词

CNN，深度学习，信息隐藏，大容量

Image Information Hiding Algorithm Based on Encoder-Decoder Network

Yajing Li, Haiyang Ding*, Tao Xiong

Department of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: Oct. 8th, 2024; accepted: Nov. 7th, 2024; published: Nov. 20th, 2024

Abstract

Traditional image steganography methods often focus on securely embedding hidden information into cover images, while paying little attention to the payload capacity. To address the issue of low embedding capacity in conventional steganography, this paper combines deep learning with image information hiding techniques. Experimental results show that the proposed algorithm achieves an embedding capacity of 24 bpp, making it one of the highest-capacity image steganography algorithms to date. Despite the large embedding capacity, the stego-images generated by the algorithm and the extracted secret images outperform other similar algorithms in both subjective visual quality and objective visual metrics such as Peak Signal-to-Noise Ratio (PSNR). This demonstrates the

*通讯作者。

文章引用：李雅静，丁海洋，熊涛. 基于编码-解码网络图像信息隐藏算法[J]. 人工智能与机器人研究, 2024, 13(4): 765-771. DOI: 10.12677/airr.2024.134078

overall superiority of the designed end-to-end steganography network.

Keywords

CNN, Deep Learning, Information Hiding, High Capacity

1. 引言

随着信息技术和数字媒体的迅猛发展，信息安全问题变得尤为重要。特别是在数字通信和互联网广泛应用的时代，信息的安全传输和隐秘通信成为了许多领域亟需解决的难题[1]。图像信息隐藏技术作为一种重要的信息保护手段，通过将秘密信息嵌入到普通图像中，使得信息传输过程更加隐蔽且难以察觉。这项技术被广泛应用于版权保护、军事通信和隐私保护等领域。

传统的图像隐写方法如最低有效位(LSB)替换、像素值差分(PVD)以及离散小波变换(DWT)，通过对载体图像的像素值进行修改，将秘密信息隐藏其中。然而，这些方法的嵌入容量有限，信息隐藏量的增加可能会导致图像质量的下降，进而增加秘密信息被检测和分析的风险。随着隐写分析技术的进步，传统方法在信息安全性和隐蔽性方面的缺陷愈发明显[2]-[7]。

近年来，深度学习技术在计算机视觉领域取得了突破性进展，为图像处理和信息隐藏提供了新的可能性。深度学习算法通过自动学习图像的特征，能够更加有效地实现高容量的信息隐藏，并在保持图像视觉质量的同时增强信息的安全性和鲁棒性。自编码器、生成对抗网络(GANs)等深度学习模型已经成功应用于图像隐写领域[8]-[10]，与传统方法相比，它们在隐写容量、隐蔽性、图像质量和抗隐写分析攻击的能力上都有显著提升。基于深度学习的图像信息隐藏技术相较于传统方法在隐写容量、安全性和图像质量上都有显著提升，尤其是在面对隐写分析攻击时表现出更强的鲁棒性。

然而，尽管这些方法取得了显著进展，但仍有改进空间，如在嵌入容量和图像质量之间的平衡、隐写分析攻击的抵抗力等方面。本文在此基础上，提出了一种基于编解码网络的更高效的大容量鲁棒图像隐写术方案，与其他隐写术方案的容量相比，本文方案在 256×256 像素彩色图像上实现了高达 24 bpp 的隐写容量。

2. 相关工作

图像信息隐藏技术已经在信息安全领域得到了广泛的应用。传统的隐写方法主要通过对载体图像的像素值进行修改来嵌入秘密信息，这些方法包括最低有效位(LSB)替换、像素值差分(PVD)以及离散小波变换(DWT)等[11]。这些技术在图像质量和嵌入容量之间取得了一定的平衡，但它们的嵌入容量有限，尤其是在面临隐写分析攻击时，其安全性和鲁棒性较差。因此，如何提高隐写容量和隐蔽性是传统隐写方法的一大挑战。

为了应对传统方法的局限性，近年来，深度学习在图像处理领域的成功应用为图像信息隐藏技术带来了新的发展方向[12]。深度学习通过其强大的特征提取能力，能够在更高的嵌入容量下保持载体图像的视觉质量，同时提高信息隐藏的安全性。

近年来，许多基于深度学习的隐写方法被提出。自编码器是其中的一种典型方法，它通过自动学习

输入图像的紧凑表示，将秘密信息嵌入到载体图像中，并通过解码器实现信息的提取。Baluja [13]等人提出了一种基于卷积神经网络(CNN)的隐写算法，该算法利用自编码器架构，在保持图像质量的同时，实现了较高的嵌入容量。

生成对抗网络(GAN)近年来在隐写领域也引起了广泛关注[14] [15]。GAN 通过生成器和判别器之间的博弈学习，不仅可以生成高质量的载体图像，还能提高对隐写攻击的抵抗能力。Zhu [16]等人提出的基于 GAN 的隐写算法，通过对抗学习，增强了隐写图像的隐蔽性，使其更难以被隐写分析检测到。

可逆隐写技术是近年来的另一个研究热点。该技术允许在提取秘密信息的同时完全恢复载体图像。传统隐写方法往往会对载体图像造成不可逆的损坏，而可逆隐写技术能够有效解决这一问题，特别是在需要高保真恢复载体图像的应用场景中[17] [18]。基于深度学习的可逆隐写方法，如通过 CNN 实现的可逆隐写，进一步提高了信息隐藏的鲁棒性和安全性。

3. 算法设计

所提出方法的工作流程如图 1 所示，由三个模块组成：预处理模块，嵌入网络和提取网络。预处理模块为嵌入网络生成载密图像做了准备。嵌入网络的目的是生成载密图像，将秘密图像隐藏在载体图像中。提取网络从载密图像中提取恢复秘密图像。预处理模块与嵌入网络一起放置在发送端，来生成隐写图像。提取网络在接收端，用于从载密图像中提取秘密图像。

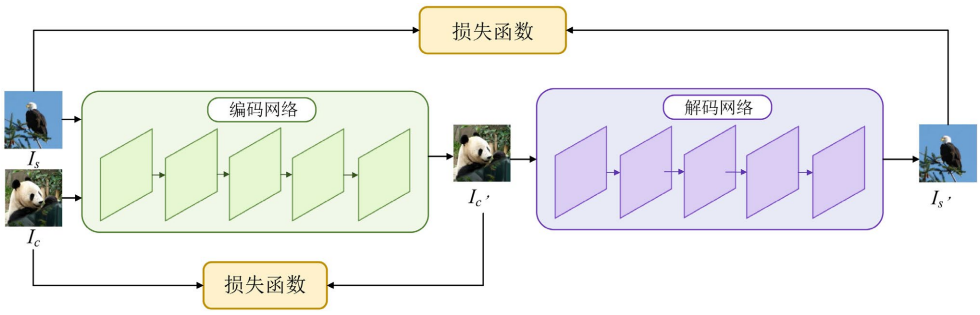

Figure 1. Overall flow chart
图 1. 总体流程图

3.1. 网络结构

编码器和解码器采用全卷积网络架构。编码器的输入是一个 256 × 256 × 6 的矩阵，包含一个 256 × 256 × 3 的 RGB 载体图像和一个同样大小的秘密图像，将这两者进行维度拼接。编码器的输出为一个 256 × 256 × 3 的 RGB 载密图像，如图 2、图 3 所示。编码器的主要任务是将秘密图像嵌入到载体图像中，同时确保生成的载密图像在视觉上保持高质量。

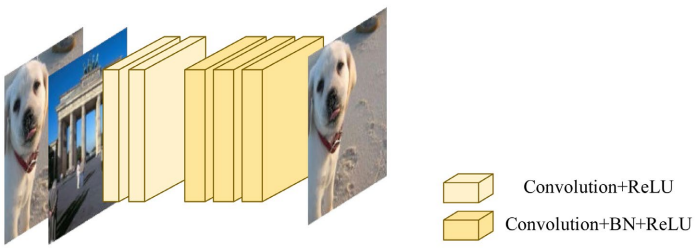

Figure 2. Encoding network structure
图 2. 编码网络结构

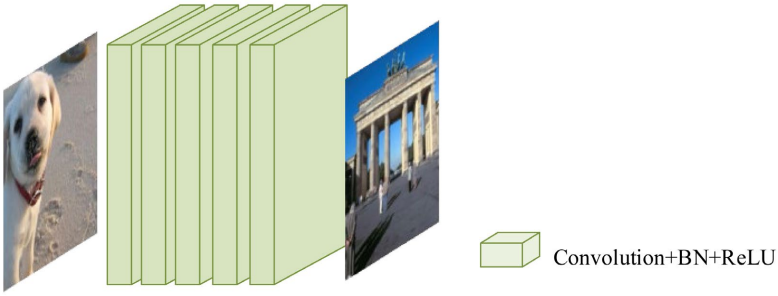

Figure 3. Decoding network structure
图 3. 解码网络结构

在嵌入秘密图像的过程中，必须考虑到解码器提取秘密图像的准确性。编码器不仅需要确保秘密图像在视觉上难以察觉，还要保证嵌入的秘密信息能够被高精度地提取和恢复。因此，嵌入网络是整个架构的关键部分。它并非只是卷积层的简单叠加，而是结合了预处理和多维度拼接等技术。预处理用于提取输入图像中的关键特征，而多维拼接则用于对封面图像和秘密图像的特征进行有效组合，确保信息嵌入的有效性和图像质量的平衡。

3.2. 损失函数

与传统的图像重建不同，图像隐写过程需要输入和输出的过程。因此，引入了自定义的损失函数以提高架构的性能。需要计算两种损失：嵌入损失 L_m 和提取损失 L_e，总损失是嵌入损失和提取损失的加权总和。公式如下：

$$L_m = \mathrm{MSE}(C, C') = \frac{\sum_{i=1}^{C}\sum_{j=1}^{H}\sum_{k=1}^{W}\left(C_{(i,j,k)} - C'_{(i,j,k)}\right)^2}{C \times H \times W} \tag{1}$$

$$L_e = \mathrm{MSE}(S, S') = \frac{\sum_{i=1}^{C}\sum_{j=1}^{H}\sum_{k=1}^{W}\left(S_{(i,j,k)} - S'_{(i,j,k)}\right)^2}{C \times H \times W} \tag{2}$$

其中，C、H 和 W 分别表示图像的通道数、高度和宽度。$I_{(i,j,k)}$ 表示图像第 i 行、第 j 列和第 k 个通道中 0 至 255 范围内的像素值。因此，模型的损失函数可以用公式表示：

$$L = L_m + \alpha \times L_e \tag{3}$$

4. 实验分析

实验在 RTX3090 GPU 上使用 Python 3.8 编程语言和 Pytorch 1.11 框架进行训练，学习率为 0.001。从 ImageNet 数据集中随机选取了 20,000 张图像作为训练集，测试集包括从 ImageNet 数据集中随机选取的另外 5000 张图像，对于每张封面图像，使用的秘密图像是从训练集或测试集中随机选取的，所有图像都预处理为 256 × 256 × 3 RBG 图像。该网络已迭代 500 次，$\alpha = 0.75$ 。

4.1. 容量分析

图像水印算法的嵌入容量是评估图像水印算法的重要标准之一，其指在不造成视觉失真的情况下，所能够嵌入的最大信息量。相对嵌入容量是本章算法衡量嵌入容量的指标。本文提出的方案将彩色图像隐藏在彩色图像中，与传统的数据隐藏方案相比有了很大的改进。为了进一步研究该方案的隐藏容量，分析了该方案的相对容量。计算方法如下：

$$EC = \frac{NS}{NC} \tag{4}$$

其中 NS 是秘密数据的大小，并且 NC 是载体图像的(像素数) × (通道数)。本算法的相对容量为 24 bpp。

4.2. 视觉质量分析

在隐写或信息隐藏方案中，以峰值信噪比(PSNR)和结构相似性(SSIM)指标作为评价标准的方案被广泛使用。图像质量由 PSNR、SSIM 等指标组成的评价表进行评价，PSNR 值高说明两幅图像之间的失真很小，隐藏图像的质量较好。类似地，在提取过程中，较高的 PSNR 可以指示所提取的图像的质量非常高。PSNR 计算方法如下：

$$PSNR = 10\log_{10}\left(\frac{\left(2^n - 1\right)^2}{MSE}\right) \tag{5}$$

结构相似性指数(SSIM)也被广泛应用于信息隐藏领域。SSIM 方法量化了亮度、对比度和结构三个因素的相似性。简化的 SSIM 可以通过诸如图像的方差和协方差的公式来计算。公式如下：

$$SSIM(x,y) = \frac{\left(2\mu_x\mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{6}$$

同时，通过对比载体图像和载密图像，可以发现重建的隐写图像与原始隐写图像没有明显差别，结果见表 1 和图 4 所示。

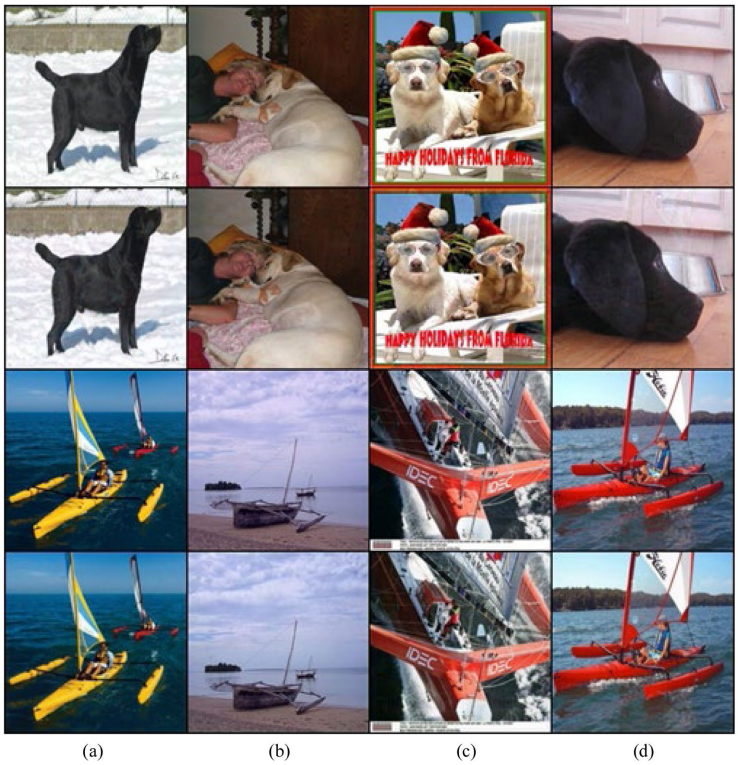

Figure 4. Images of experimental results
图 4. 实验结果图像

Table 1. PSNR and SSIM of experimental result images
表 1. 实验结果图像的 PSNR 和 SSIM

Compare image	PSNR (dB)	SSIM
Figure 3(a)	32.7088	0.8972
Figure 3(b)	33.8801	0.9182
Figure 3(c)	32.9825	0.9073
Figure 3(d)	33.5104	0.8770

5. 结论

本文提出了一种基于卷积神经网络(CNN)的端到端图像隐写方法，系统结构包括预处理模块、嵌入网络和提取网络。预处理模块对载体图像进行处理，以便嵌入网络能够将秘密图像嵌入其中，而提取网络则负责从生成的隐写图像中提取出秘密图像。实验结果表明，与传统隐写方法及其他深度学习隐写方法相比，该方法在峰值信噪比(PSNR)上具有显著优势，证明在信息隐藏中的安全性和鲁棒性更高。此外，该方法的容量为 24 bpp，表明在实现高容量隐写的同时，仍保持了良好的隐写效果。

基金项目

北京市教委科研计划(KM202010015009, KM202110015004)；北京市高等教育学会项目(MS2022093, MS2023204)；北京市数字教育研究课题(BDEC2023619095)；北京印刷学院思政重点项目(20240053)；北京印刷学院青年卓越项目(Ea202411)。

参考文献

[1] Xian, Y., Wang, X., Zhang, Y., Wang, X. and Du, X. (2021) Fractal Sorting Vector-Based Least Significant Bit Chaotic Permutation for Image Encryption. *Chinese Physics B*, **30**, Article ID: 060508. https://doi.org/10.1088/1674-1056/abda35

[2] Duan, X., Jia, K., Li, B., Guo, D., Zhang, E. and Qin, C. (2019) Reversible Image Steganography Scheme Based on a U-Net Structure. *IEEE Access*, **7**, 9314-9323. https://doi.org/10.1109/access.2019.2891247

[3] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI* 2015, Springer International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[4] Zhang, R., Dong, S. and Liu, J. (2018) Invisible Steganography via Generative Adversarial Networks. *Multimedia Tools and Applications*, **78**, 8559-8575. https://doi.org/10.1007/s11042-018-6951-z

[5] Hayes, J. and Danezis, G. (2017) Generating Steganographic Images via Adversarial Training. *Proceedings of the* 31*st Annual Conference on Neural Information Processing Systems*, La Jolla, 1955-1964.

[6] Hu, D., Wang, L., Jiang, W., Zheng, S. and Li, B. (2018) A Novel Image Steganography Method via Deep Convolutional Generative Adversarial Networks. *IEEE Access*, **6**, 38303-38314. https://doi.org/10.1109/access.2018.2852771

[7] Agustsson, E. and Timofte, R. (2017) NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Honolulu, 21-26 July 2017, 1122-1131. https://doi.org/10.1109/cvprw.2017.150

[8] Ardizzone, L., Kruse, J., Rother, C. and Kothe, U. (2018) Analyzing Inverse Problems with Invertible Neural Networks. arXiv: 1808.04730.

[9] Baluja, S. (2017) Hiding Images in Plain Sight: Deep Steganography. In: Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R., Eds., *Advances in Neural Information Processing Systems* 30, (*NeurIPS*).

[10] Baluja, S. (2019) Hiding Images within Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 1685-1697.

[11] Dinh, L., Sohl-Dickstein, J. and Bengio, S. (2016) Density Estimation Using Real NVP. arXiv: 1605.08803.

[12] Gilbert, A., Zhang, Y., Lee, K., Zhang, Y. and Lee, H. (2017) Towards Understanding the Invertibility of Convolutional Neural Networks. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, 19-25 August 2017, 1703-1710. https://doi.org/10.24963/ijcai.2017/236

[13] Baluja, S. (2017) Hiding Images in Plain Sight: Deep Steganography. In: Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R., Eds., *Advances in Neural Information Processing Systems* 30, (*NeurIPS*).

[14] He, J.W., Dong, C. and Qiao, Y. (2020) Interactive Multi-Dimension Modulation with Dynamic Controllable Residual Learning for Image Restoration. arXiv: 1912.05293.

[15] Hetzl, S. and Mutzel, P. (2005) A Graph-Theoretic Approach to Steganography. In: Dittmann, J., Katzenbeisser, S. AND Uhl, A., Eds., *Communications and Multimedia Security*, Springer Berlin Heidelberg, 119-128. https://doi.org/10.1007/11552055_12

[16] Zhu, J., Kaplan, R., Johnson, J. and Fei-Fei, L. (2018) HiDDeN: Hiding Data with Deep Networks. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV* 2018, Springer International Publishing, 682-697. https://doi.org/10.1007/978-3-030-01267-0_40

[17] Ho, J., Chen, X., Srinivas, A., *et al*. (2019) Flow++: Improving Flow-Based Generative Models with Variational Dequantization and Architecture Design. *International Conference on Machine Learning* (ICML), 3.

[18] Huang, C.W., Krueger, D., Lacoste, A. and Courville, A. (2018) Neural Autoregressive Flows. arXiv: 1804.00779.