

基于对抗蒸馏的抗视频编码视频隐藏技术

汪红松¹, 李享运¹, 曾碧卿²

(1. 华南师范大学人工智能学院; 2. 华南师范大学阿伯丁数据科学与人工智能学院, 广东 佛山 528225)

摘要: 现有的深度视频隐藏技术在网络训练阶段, 采用预训练的固定参数视频编码模拟网络替代真实视频编码器, 易导致视频隐藏网络过拟合, 进而降低其对真实视频编码器的抵抗能力。针对此问题, 提出了一种基于对抗蒸馏的抗视频编码的视频隐藏模型。首先, 设计一个基于三维Unet的视频编码模拟器, 以更真实地模拟视频编码器; 其次, 通过使用三维可逆神经网络构建视频隐藏模型, 从而提高其不可见性; 最后, 采用对抗蒸馏训练方式, 通过交替训练视频隐藏网络和编码模拟器, 解决了视频隐藏网络训练中的过拟合问题。在 REDS 数据集的测试中, 该模型在 CRF 为 10 时提取秘密视频的 PSNR 和 SSIM 分别提高了 1.78 dB 和 0.019 25; 在 CRF 为 16 时提取秘密视频的 PSNR 和 SSIM 分别提高了 2.815 dB 和 0.027 5。

关键词: 视频隐藏; 抗视频编码; 生成对抗网络; 知识蒸馏

DOI: 10.11907/rjdk.251178

中图分类号: TP309

文献标识码: A

文章编号: 1672-7800(2026)005-0116-10

扫描二维码阅读全文:



Video Hiding Technology Based on Adversarial Distillation against Video Encoding

WANG Hongsong¹, LI Xiangyun¹, ZENG Biqing²

(1. School of Artificial Intelligence, South China Normal University;

2. Aberdeen Institute of Data Science and Artificial Intelligence, South China Normal University, Foshan 528225, China)

Abstract: Some deep video hiding technologies use pre-trained fixed parameter video network to simulate the analog video encoder in the network training phase, resulting in over fitting of the video hiding network, which reduces the resistance of the video hiding network to the real video encoder. To solve these problems, this paper proposes an anti video encoding video hiding technology via adversarial distillation. Firstly, a 3D-Unet video coding simulator is designed to simulate the video encoder more realistically; Secondly, a 3D invertible neural network is used to construct a video hiding model to improve the invisibility of the video hiding model; Finally, the adversarial distillation training method is used to alternately train the video hiding network and the video coding simulator to solve the problem of over fitting in the video hiding network training. In the test on the REDS dataset, when the CRF is 10, the PSNR and SSIM of the secret video extracted by the proposed model are increased by 1.78 dB and 0.019 25 respectively; when the CRF is 16, the PSNR and SSIM of the extracted secret video are increased by 2.815 dB and 0.027 5 respectively.

Key Words: video hiding; anti-video encoding; generative adversarial networks; knowledge distillation

0 引言

视频隐藏技术是一种通过将秘密视频嵌入到载体视频中生成载密视频的方法, 使得监听方无法感知载体中是否含有秘密视频, 从而达到隐蔽传输秘密视频的目的^[1-2]。

视频隐藏技术在网络传输、军事和医疗信息保护等方面都有着非常广泛的应用。随着深度学习技术的发展, 目前的深度视频隐写技术主要使用自编码器—解码器结构或可逆神经网络来实现秘密视频的嵌入和提取, 相比于传统视频隐写技术提高了嵌入容量和不可见性^[3-5]。

然而, 视频编码器在实际应用中对深度视频隐藏技术

收稿日期: 2025-04-02

基金项目: 广东省基础与应用基础研究基金项目(2022A1515140110)

作者简介: 汪红松(1975-), 男, 博士, 华南师范大学人工智能学院讲师、硕士生导师, 研究方向为自然语言处理、情感分析、对话系统; 李享运(1998-), 男, 华南师范大学人工智能学院硕士研究生, 研究方向为隐写术、信息隐藏技术; 曾碧卿(1969-), 男, 博士, 华南师范大学阿伯丁数据科学与人工智能学院教授、硕士生导师, 研究方向为自然语言处理、人工智能、大数据。本文通讯作者: 曾碧卿。

构成了艰巨挑战。传统视频隐藏技术通常在视频编码的冗余位置嵌入秘密信息,而深度视频隐写技术往往将视频解码后送入神经网络进行处理,随后将输出结果编码为载密视频。在视频编码过程中容易造成秘密视频的丢失,从而导致无法从载密视频中恢复清晰的秘密视频。现有的深度视频隐藏技术在网络训练阶段,采用预训练的固定参数视频编码模拟网络替代真实视频编码器,易导致视频隐藏网络过拟合,进而降低其对真实视频编码器的抵抗能力。因此,如何提高深度视频隐藏技术的抗视频编码能力,尽可能地发挥深度视频隐藏技术的高嵌入容量、高不可见性和低失真度优势,成为目前视频隐藏技术领域所关注的重点问题。

1 相关研究

1.1 传统视频隐藏技术

传统视频隐藏技术可分为编码域与非编码域。非编码域的视频隐藏技术是将视频拆分为连续的图像集合,通过选取其中不重要的视频帧图像直接修改,从而降低对载体视频的影响。为了获得更好的鲁棒性和不可见性,许多非编码域视频隐藏算法更倾向于在变换域中修改图像,如使用离散小波变换(DWT)或离散余弦变换(DCT),在变换域中嵌入秘密信息,尽可能地降低秘密信息的肉眼可见程度。编码域的方法则是针对特定的视频编码器,如广泛使用的H.264/AVC或H.265/HEVC。这种方法通过基于熵值的检测来分析视频压缩中的运动向量,从而选择适合的位置嵌入秘密信息^[6]。Bai等^[7]提出了一种基于编码块大小和自适应变换隐藏秘密信息的算法,在H.265/HEVC视频编码下的表现非常出色。而Liu等^[8]则通过结合3种帧内预测与离散正弦变换的系数,有效防止了帧内失真漂移,获得了较强的鲁棒性。

1.2 深度视频隐藏技术

随着深度学习技术的发展,学者们开始尝试使用深度神经网络构建图像隐藏模型^[9-12]。Zhu等^[13]提出了HiD-DeN算法,使用自编码器-解码器结构的神经网络将二值信息嵌入到图像中。BaLuja^[14]则使用卷积神经网络,将一张图像嵌入到同样大小的图像当中。在深度图像隐写技术研究基础上,Weng等^[15]提出了将视频隐藏到视频的视频隐藏模型,RiViE^[16]使用时空生成器将秘密信息嵌入到视频中。Jaiswal等^[17]则在两者的基础上,提出了一种利用3D卷积神经网络的视频隐藏新框架,其模型将视频分割成连续的8帧图像序列作为输入,使用3D卷积神经网络提取相邻帧数之间的关系,生成载密视频。自编码器-解码器的视频隐藏技术相较于传统的视频隐藏技术具有嵌入容量大的优势。然而,该类视频隐藏技术的不可见性和恢复秘密视频能力较弱,其生成的载密视频通常会存在秘密视频的伪影,容易被肉眼观测。同时,恢复的秘密视频存

在一定程度的失真,无法清晰准确地辨认其中细节信息。

随着可逆神经网络技术广泛应用于超分辨率重建和图像着色领域,许多学者尝试使用可逆神经网络构建图像或视频隐写模型^[18-20]。Lu等^[21]首先使用了可逆神经网络构建图像隐写模型,实现了在单张彩色图像中嵌入高达5张同尺寸的秘密图像。Jing等^[22]则提出了Hinet,结合离散小波变换和可逆神经网络构建图像隐写模型,具有更高的不可见性和更小的秘密图像失真度。Mou等^[23]则提出了LF-VSN框架,通过使用滑动窗口和可逆神经网络,实现了在单一视频中嵌入多个等长视频。

上述基于深度学习的视频隐藏网络未考虑视频编码引起的失真,导致经视频编码压缩后,无法从解码视频中成功恢复秘密信息。然而,部分视频水印研究已将视频编码的影响纳入考量。Zhang等^[24]在训练过程中使用预训练的3D卷积神经网络,通过卷积神经网络模拟非可微的视频编码器,使得深度视频水印能够成功学习视频编码器带来的影响,同时,添加了一种增强H264/AVC压缩鲁棒性的深度视频水印,使用光流模块结合Jpeg压缩以模拟视频编码器。而在视频隐藏领域,刘勇等^[25]则提出了一种无帧内失真漂移的视频隐写算法。Shen等^[26]提出了视频编码模拟器,首次在使用视频编码模拟器中模拟了视频编码带来的影响。通过预训练视频编码模拟器,并将其加入到视频隐藏模型的训练过程中,显著提升了隐藏模型抗编码能力。Yang等^[27]混合了Jpeg压缩、2D高斯模糊和3D高斯模糊,构建了视频编码模拟层,在低嵌入率的环境下获得了较好的效果。然而,上述方法中使用的视频编码模拟器都是静态的,随着视频隐藏模型训练的深入,模型往往会试图欺骗这些模拟器,导致过拟合。这种过拟合会使模型对真实视频编码器的抵抗能力下降。在实际应用中,虽然模型输出的张量已编码为载密视频,但从载密视频中提取的秘密视频却无法清晰识别。

目前深度视频隐藏技术在不可见性、嵌入容量和失真程度上相比于传统视频隐藏技术有着更好的表现。尤其是基于可逆神经网络实现的视频隐藏模型,在载密视频不进行编码的情况下可以嵌入4~5段同等长度和分辨率的秘密视频。然而,深度视频隐藏模型受到视频编码器的影响,其从编码后的载密视频中恢复秘密视频的能力较为有限,这使得深度视频隐藏技术在实际应用中难以得到广泛应用。

为了解决以上问题,本文提出一种基于对抗蒸馏的视频隐藏模型,进一步提升了视频隐藏模型抵抗视频编码的能力。该模型首先设计了一个三维的Unet,作为视频编码模拟器,其在时间和空间维度上的模拟能力相比基于Jpeg压缩和高斯模糊混合的模拟器更为优秀,且具有更强的泛用性。其次,该模型引入了一种对抗蒸馏的训练策略,结合了生成对抗网络(GAN)与知识蒸馏的核心优势:在训练视频隐藏网络时,利用视频编码模拟器对其输出的载密视

频进行处理;而在训练视频编码模拟器时,则固定视频隐藏网络的参数,通过知识蒸馏使模拟器能够学习视频编码对视频隐藏网络的实际影响。在这种交替训练机制下,视频编码模拟器的模拟精度与视频隐藏网络的抗视频编码能力可逐步提升。本文的主要贡献如下:

(1)设计了三维可逆视频隐藏网络和三维Unet视频编码模拟器,相比于现有的网络有着更好的视频隐藏性能和视频编码模拟能力。

(2)在视频隐写模型中引入了对抗蒸馏训练方式,通过结合生成对抗网络和知识蒸馏,提高了视频隐藏模型抗视频编码的能力。

(3)实验证明,本文提出的视频隐藏模型所生成的载密视频具有更好的抗视频编码能力。载密图像和恢复秘密图像的PSNR分别提高了0.39 dB和0.32 dB,在SSIM相似度对比上分别提高了1.5%和0.5%。

2 基于对抗蒸馏的抗视频编码的视频隐藏技术

本文提出的视频隐藏框架的目的是将一个视频隐藏

到另一个视频当中,并且在经过视频编码后仍然可以从载密视频中恢复秘密视频。与其他视频隐藏模型相比,本文不仅将视频当作连续的图像序列,还将对隐藏网络的输出进行视频编码。因此,本文提出了一种基于对抗蒸馏的抗视频编码的视频隐写模型。该模型由视频隐藏提取网络和视频编码模拟器两个部分组成。其中,视频隐藏提取网络使用可逆神经网络实现。可逆神经网络的正向和反向过程,分别对应着视频隐藏和提取秘密视频这两个功能。由于可逆神经网络的网络参数在正向和反向过程中是共享的,因此,其在视频隐藏方面相较于编码器-解码器网络具有更好的性能。此外,本文设计了一种基于3D-Unet的视频编码模拟器,用于模拟视频编码对载密视频的影响。本文采用对抗蒸馏的方式来训练视频隐藏网络和视频编码模拟器,使得视频编码模拟器的模拟能够更加接近实际的视频编码,从而在尽可能的情况下对载密视频中的隐藏秘密视频进行攻击。

2.1 视频隐藏与提取网络

本文使用了可逆神经网络用于秘密视频的隐藏和提取,其网络结构和流程图如图1所示。

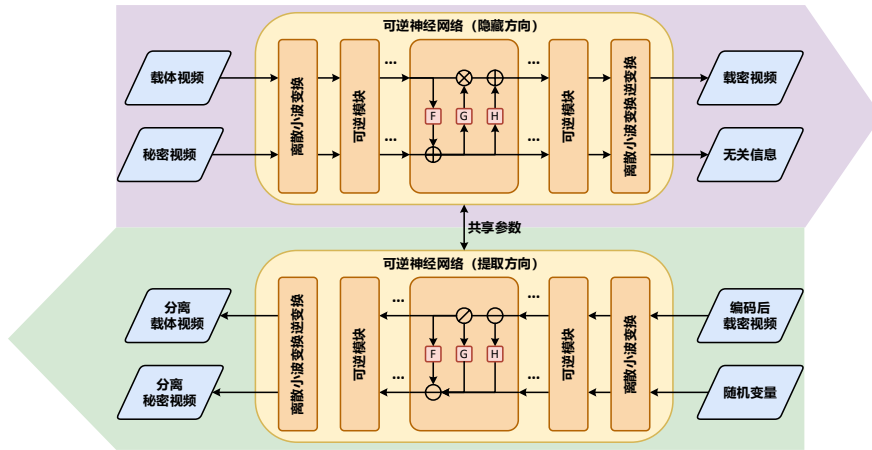


Fig. 1 Overall structure and process of invertible neural network

图1 可逆神经网络结构和流程图

视频隐藏与提取网络由离散小波变换模块和若干可逆子模块构成。其中,离散小波变换模块采用 Haar 离散小波变换;每个可逆子模块包含 3 个仿射函数(F、G、R),用于建立输入端与输出端的映射关系。需要说明的是,离散小波变换模块可提升隐写模型的视觉效果,该作用已在文献[28-31]中得到论述,本文不再展开。仿射函数 F、G、R 使用密集连接块 Dense Block 实现,用于提取载体视频 X_{cover} 和秘密视频 X_{secret} 的特征。嵌入过程与提取过程的参数共享,不需要训练额外的提取网络。

在嵌入过程中,可逆隐写模块输入端为秘密视频 X_{secret} 和载体视频 X_{cover} ,输出端为载密视频 X_{stego} 和损失信息 R 。其计算流程如公式(1)所示。

$$\begin{aligned} X_{cover}^{i+1} &= X_{cover}^i + F(X_{secret}^i) \\ X_{secret}^{i+1} &= X_{secret}^i \otimes e^{a(G(X_{cover}^{i+1}))} + R(X_{cover}^{i+1}) \end{aligned} \quad (1)$$

在嵌入流程中,将秘密视频 X_{secret} 和载体视频 X_{cover} 经过离散小波变换后输入可逆隐写模块。通过若干个可逆子模块处理后,再经过离散小波逆变换,最终输出损失信息 R 与载密视频 X_{stego} 。其中,损失信息 R 属于随机高斯噪声 Z 。

在提取过程中,可逆隐写模块输入端为载密视频 X_{stego} 和随机采样的高斯噪声 z ,输出端为秘密视频 X_{rev_secret} 和分离的载体视频 X_{rev_cover} ,其计算流程如公式(2)所示。

$$\begin{aligned} X_{rev_secret}^i &= (X_{rev_secret}^{i+1} - R(X_{rev_cover}^{i+1})) \otimes e^{-a(G(X_{rev_cover}^{i+1}))} \\ X_{rev_cover}^i &= X_{rev_cover}^{i+1} - F(X_{rev_secret}^i) \end{aligned} \quad (2)$$

2.2 基于Unet的视频编码模拟器

真实的视频编码是不可微分的模块,无法直接加入到视频隐藏和提取模块的训练中。为了使视频隐藏和提取

模块能够有效抵抗真实的视频编码,本文使用基于Unet的网络结构来模拟视频编码,如图2所示。这些模拟网络

在训练过程中用于攻击载密视频,从而增强模型的鲁棒性。

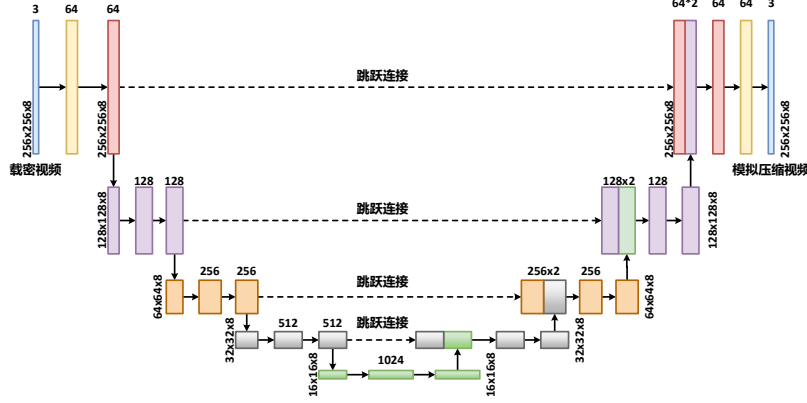


Fig. 2 Overall structure and process of 3D-Unet video encoding simulator

图2 3D-Unet视频编码模拟器流程和结构图

视频编码模拟器基础为U-net架构,由编码-解码2个部分组成,同时在解码过程中加入了跳跃连接模块。编码阶段由4次下采样双重三维卷积层实现,在2次卷积层之间加入了池化层。此外,在解码阶段,为了减小视频的低频信息,本文加入了跳跃连接用以保留视频的低频信息,并使用反卷积、上采样和双卷积等方法来恢复视频。

2.3 训练流程与损失函数

以往工作中,视频编码模拟器通常在预训练后固定其参数,然后在视频隐藏网络的训练过程中对载密视频进行攻击。这种做法可能导致视频隐藏网络更倾向于欺骗视频编码模拟器,从而使其对真实视频编码的抵抗能力变弱。为解决上述问题,本文设计基于对抗蒸馏的训练流程,核心是让视频编码模拟器与视频隐藏网络以交替方式开展训练。具体而言,在训练视频隐藏和提取模块时,会固定视频编码模拟器的参数,确保模块能在稳定的模拟编码条件下学习抗干扰能力。在训练视频编码模拟器时,则会固定视频隐藏和提取模块,确保视频编码模拟器可以更好地模拟视频编码对隐写模型造成的影响。

图3为视频隐藏网络训练流程图。具体流程如下:首先,载体视频和秘密视频通过视频隐藏模块处理生成载密视频;其次,该载密视频被输入到视频编码模拟器中进行攻击,攻击后的载密视频随后被输入到视频提取模块;最后,从中提取出秘密视频。

在视频隐藏网络的训练过程中,隐藏模型关注载体视频和载密视频之间的相似度,同时也关注秘密视频与提取后的秘密视频之间的相似度。为此,优化器使用均方差来优化这两个相似度,其计算流程如公式(3)所示。

$$\mathcal{L}_{Con} = (X_{cover} - X_{stego})^2 \quad (3)$$

$$\mathcal{L}_{Rev} = (X_{secret} - X_{rev_secret})^2$$

整体损失函数 $\mathcal{L}_H$ 如公式(4)所示。

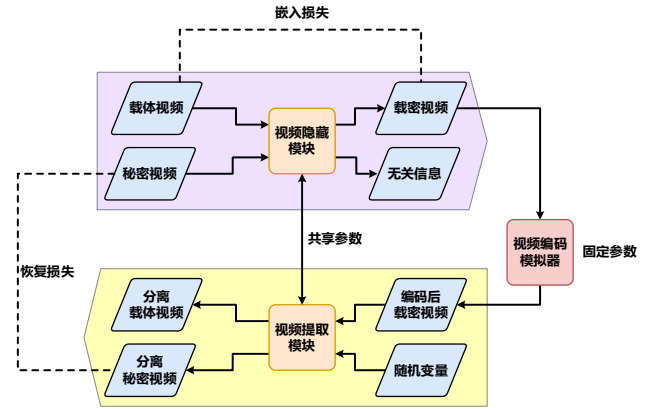


Fig. 3 The training process of video hiding network

图3 视频隐藏网络训练流程

$$\mathcal{L}_H = \mathcal{L}_{Con} + \mathcal{L}_{Rev} \quad (4)$$

图4为3D-Unet视频编码模拟器训练流程图。具体流程如下:首先,训练视频编码模拟器时先固定视频隐藏与提取模块的参数;其次,将载体视频与秘密视频输入参数固定的视频隐藏模块,处理后输出载密视频;然后,用真实视频编码器对载密视频进行双路径处理,得到真实编码载密视频,随后用待训练的视频编码模拟器处理得到模拟编码载密视频;最后,将两种编码载密视频分别输入视频提取模块,提取后得到真实提取秘密视频与模拟提取秘密视频。

在视频编码模拟器训练流程中,优化器使用损失函数 $\mathcal{L}_S$ 来优化视频隐藏和提取模块,如公式(5)所示。

$$\mathcal{L}_{Com} = (X_{stego} - X_{compress_stego})^2 \quad (5)$$

$$\mathcal{L}_{Rev} = (X_{secret} - X_{rev_secret})^2$$

其中, $\mathcal{L}_{Com}$ 为真实编码载密视频和模拟编码载密视频的均方差, $\mathcal{L}_{Rev}$ 为真实提取秘密视频和模拟提取秘密视频的均方差。对于整体的视频隐藏模型而言,更加关注视频

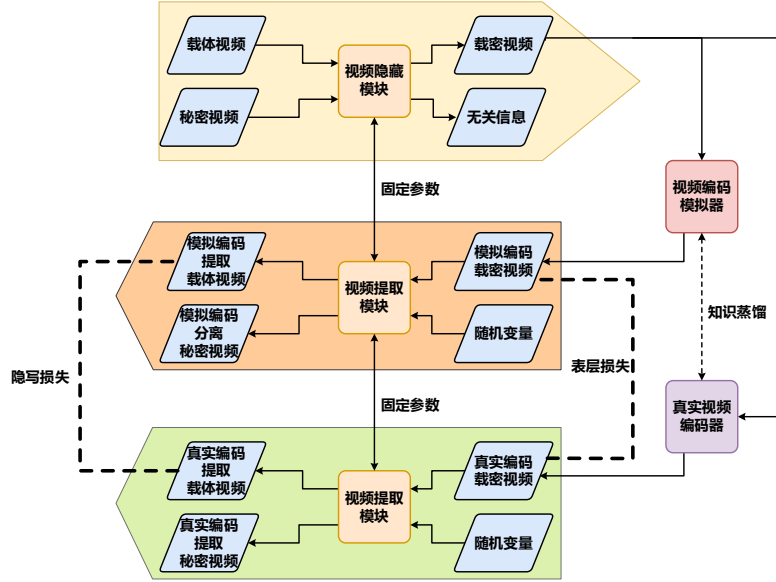


Fig. 4 The training process of 3D-Unet video encoding simulator

图 4 3D-Unet 视频编码模拟器训练流程

编码模拟器对载密视频中的秘密视频的影响程度,因此 $\mathcal{L}_{Com}$ 和 $\mathcal{L}_{Rev}$ 的比例为 1:5, 整体损失函数 $\mathcal{L}_S$ 如公式(6)所示。

$$\mathcal{L}_S = 0.2 * \mathcal{L}_{Com} + \mathcal{L}_{Rev} \quad (6)$$

3 实验与分析

3.1 实验设置与评价指标

本文提出的模型采用固定划分的训练集与测试集,分别使用 REDS 和 UCF101 数据集训练模型,并在测试结果中单独标记两种数据集训练模型的表现。其中,REDS 数据集包含 270 条具有时间连续性的图像序列,每条序列含 100 帧 1 280×720 分辨率图像,240 条序列用于模型训练,30 条用于测试。训练阶段随机裁剪 8 帧 256×256 分辨率的图像序列块作为载密视频与秘密视频的输入,测试阶段则选取序列中间 96 帧图像进行嵌入测试。UCF101 数据集包含 13 320 条已通过 H.264/AVC 编码的 320×240 分辨率视频,本文参考 VHNet 的实验设置,选用其中 8 000 条用于训练,另取 2 660 条作为验证集,300 条作为测试集。实验基于 PyTorch1.12.1 框架搭建,使用 4 张 NVIDIA Tesla T4 GPU 训练,优化器采用 Adam(参数 $\beta_1=0.9$ 、 $\beta_2=0.99$),初始学习率设为 1×10^{-4} ,训练批次大小(batch size)设定为 8。

在视频隐藏领域,重点关注的两个关键问题是视频对之间的失真程度。首先,关注载密视频与载体视频之间的失真程度;其次,关注提取的秘密视频与原始秘密视频之间的失真程度。本文使用峰值信噪比 PSNR(Peak Signal to Noise Ratio)和结构相似性 SSIM(Structural Similarity)对视频的质量进行评估,其计算方法如公式(7)所示。

$$\begin{aligned} MSE(X, Y) &= \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [X(i, j) - Y(i, j)]^2 \\ PSNR(X, Y) &= 20 * \log \frac{MAX_i}{\sqrt{MSE(X, Y)}} \\ SSIM(X, Y) &= \frac{(2\mu_x \mu_y + c_1)(\sigma_x + \sigma_y + \mu_{c_2})}{(\mu_x^2 + \mu_y^2 + \mu_x)(\sigma_x^2 + \sigma_y^2 + \mu_x)} \end{aligned} \quad (7)$$

在 PSNR 公式中, X 表示真实视频, Y 表示模型生成视频, MAX_i 表示像素点可取的最大值, n, m 分别为视频的最大宽和高。

在 SSIM 公式中, μ_x 和 μ_y 分别表示视频 X 和视频 Y 之间的均值, σ_x 和 σ_y 分别表示视频 X 和视频 Y 的标准差, c_1 与 c_2 均为维持稳定的常数。当视频像素值最大值为 255 时, $c_1 = 6.5025$, $c_2 = 58.5225$ 。

其中, PSNR 指标越高则表示视频的失真程度越低, SSIM 越接近 1.0 则表示两幅视频之间的结构相似性最高。

3.2 不可见性和失真程度对比实验

本文选取了目前常见的视频隐写模型 HiDDeN、LF-VSN、VStegNet、VHNet、DRANet^[32]和 StegaNeRV^[33]作对比。其中 HiDDeN 由于其性能较弱,平均每像素点仅能嵌入 0.4bit 的秘密信息,因此本文使用卷积神经网络对 HiDDeN 进行改进,使其可以嵌入同等尺寸的秘密视频。VStegNet、VHNet 本文选择了其在 UCF101 上训练后的模型;而 LF-VSN 本文选择了其在 REDS 数据集训练的最终模型。在视频编码方面,本文选取了目前常用的 H.264/AVC 视频编码,在 CRF 为 10 或 16 的情况下分别对生成的载密视频进行视频编码。视频编码的 crf 值越大,其编码造成的损失越大。在 REDS 数据集下不同视频隐写模型生成载密视频和提取秘密视频的性能对比如表 1 所示。

由表 1 可知, VHNet、VstegNet、DRANet、StegaNeRV 均

Table 1 Comparison of the performance of video steganography models on the REDS testset

表 1 REDS 测试集下视频隐写模型性能对比						
模型	载密视频		提取秘密视频(crf=10)		提取秘密视频(crf=16)	
			PSNR	SSIM	PSNR	SSIM
HiDDeN	31.75	0.877	11.68	0.382	11.71	0.379
LF-VSN	44.26	0.991	11.05	0.378	11.03	0.368
VStegNet	36.23	0.981	19.86	0.756	17.14	0.702
VHNet	33.67	0.914	31.04	0.883	27.21	0.783
DRANet	31.03	0.902	30.42	0.839	27.46	0.779
StegaNeRV	33.84	0.918	31.12	0.849	27.29	0.782
本文(REDs)	34.06	0.929	29.89	0.851	27.59	0.789
本文(UCF101)	33.86	0.929	31.06	0.858	27.62	0.791

具有视频编码抗性,相对的 LF-VSN 和 HiDDeN 对于视频编码没有抗性。LF-VSN 所生成的载密视频的质量最好,但是由于在训练过程中完全没有考虑视频编码的问题,因此,其生成的载密视频在经过视频编码后,其中的秘密信息会完全丢失。在具有视频编码抗性的模型中,VStegNet 生成载密视频的质量最好,但是其在提取秘密视频的质量最差。而本文提出的模型所生成的载密视频和经过视频编码后提取秘密视频的视频质量均较优。与具有视频编码抗性的模型的平均值进行对比,本文在 REDS 数据集上训练的模型所生成的载密视频质量在 PSNR 和 SSIM 上分别提高了 0.36 dB 和 0.000 25。当 CRF 为 10 时,在提取秘密视频的视频质量比较中,PSNR 和 SSIM 分别提高了 1.78 dB 和 0.019 25;当 CRF 为 16 时,PSNR 和 SSIM 分别提高了 2.815 和 0.027 5。

表 2 为 UCF101 测试集上视频隐写模型的性能对比。从表中可见,HiDDeN 与 LF-VSN 在该测试集上仍不具备视频编码抗性,而 VStegNet、VHNet、DRANet 及 StegaNeRV 均展现出一定的视频编码抗性,其中 StegaNeRV 生成的载密视频 PSNR 值最优,VStegNet 的 SSIM 值表现最佳。在提取秘密视频的指标对比中,本文基于 UCF101 训练集训练的模型,在 CRF=10 和 CRF=16 条件下,PSNR 与 SSIM 指标均位居第一。值得注意的是,在 UCF101 测试集上,本文基于 REDS 数据集训练的模型性能不及基于 UCF101 训练的

Table 2 Comparison of the performance of video steganography models on the UCF101 testset

表 2 UCF101 测试集上视频隐写模型的性能对比						
模型	载密视频		提取秘密视频(crf=10)		提取秘密视频(crf=16)	
			PSNR	SSIM	PSNR	SSIM
HiDDeN	27.76	0.877	11.65	0.380	11.70	0.373
LF-VSN	42.69	0.989	11.03	0.376	11.04	0.368
VStegNet	31.79	0.927	19.36	0.736	17.12	0.694
VHNet	33.71	0.924	30.13	0.853	27.01	0.773
DRANet	30.95	0.913	29.86	0.821	26.89	0.762
StegaNeRV	33.91	0.925	30.07	0.844	26.63	0.771
本文(REDs)	34.06	0.929	30.26	0.862	27.23	0.782
本文(UCF101)	33.81	0.919	30.02	0.843	26.57	0.771

模型,这一差异主要由数据集特性决定:UCF101 提供的均为经 H.264/AVC 编码后的视频,而 REDS 则是未编码的图像序列。这一结果表明,在实际应用中需根据具体传输信道特性,选择适配的模型生成载密视频进行传输,才能获得更优的秘密视频提取性能。

3.3 视觉效果对比

表 3 展示了 REDS 数据集下视频隐写模型在 crf=10 和 crf=16 时的视觉效果对比,实验采用 H.264/AVC 视频编码,分辨率为 512×512,视频总长度 96 帧,且以 8 帧为单位分段嵌入秘密视频,比特率设定为 4 000 K。从图中可观察到,训练后的视频隐写模型生成的载密视频,在 crf=10 与 crf=16 条件下视觉效果差距较小,无论是纹理复杂区域还是纹理平滑区域,均保持较好的视觉表现;但经过 H.264 视频压缩编码后,所有模型提取的秘密视频清晰度均有所下降。

从表 3 可以看出,HiDDeN 因采用图像隐藏模型的噪声层,其对抗视频压缩编码的性能最差;VStegNet 与 VHNet 虽使用相同的视频编码模拟网络,但二者提取的秘密视频均出现一定程度的伪影与模糊;而本文模型借助对抗蒸馏机制与视频编码模拟器,在提取秘密视频时表现出更优性能,且基于 REDS 数据集训练的模型,提取结果比基于 UCF101 数据集训练的模型更清晰。其原因在于 UCF101 数据集的视频存在分辨率较低、画面较模糊且已预先经过视频编码处理的问题,使得基于该数据集训练的模型在 REDS 数据集上表现较弱。

3.4 模型效率和参数量对比

表 4 展示了当前广泛使用的 VstegNet、VHNet、LF-VSN、DRANet 和本文模型在参数量、生成载密视频耗时以及提取秘密视频耗时等方面的对比分析。实验中选用了 REDS 数据集视频进行秘密信息的嵌入和提取,设定的视频分辨率为 512×512,视频总长度为 96 帧。其中,以 8 帧作为分段嵌入秘密视频,且质量因子 CRF 分别设置为 10 和 16。测试环境为:CPU i5-12600kf,内存 32 G,以及 NVIDIA RTX 2080ti,分别用于生成载密视频和提取秘密视频。

其中,VStegNet 由于使用自编码器-解码器结构分别实现秘密视频的嵌入和提取功能,导致其网络参数量巨大、运行效率较低,生成 96 帧载密视频需要耗时近 3min。DRANet 同样使用自编码器-解码器的结构,虽然参数量相比于 VStegNet 有所下降,其生成载密视频仍然需要耗时超过 2 min。而 VHNet、LF-VSN 和本文提出的模型使用可逆神经网络实现秘密信息的嵌入和提取。其中,VHNet 和 LF-VSN 使用二维可逆神经网络实现参数量较低的同时运行效率同样较高。本文提出的模型通过三维可逆神经网络实现,且包含视频编码模拟器,因此网络参数较大。然而,视频编码器仅在训练过程中与视频隐写网络进行对抗训练,并承担蒸馏视频编码器的任务,而在实际应用场景中则不运行。综上,三维可逆神经网络相比二维可逆神经网络耗时更长,但能够实现更好的时间特征融合,从而在

Table 3 Comparison of crf=10 and crf=16 visuals on the REDS dataset
表 3 REDS数据集下 crf=10 和 crf=16 视觉效果对比图

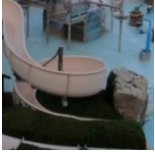
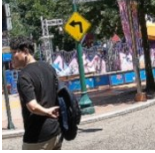
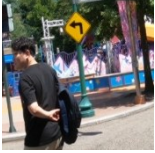
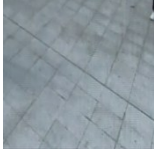
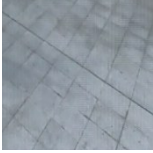
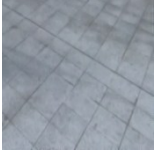
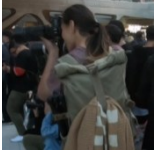
CRF	原图	HiDDeN	VStegNet	VHNet	本文(REDs)	本文(UCF101)
10	载密视频					
	秘密视频					
	载密视频					
	秘密视频					
16	载密视频					
	秘密视频					
	载密视频					
	秘密视频					

Table 4 Comparison of parameter quantity, time spent for concealing container video and revealing secret video

表 4 参数量、生成载密视频耗时和提取秘密视频耗时对比

模型	参数量/百万	生成载密视频 耗时/s	提取秘密视频 耗时/s
VstegNet	42.57	160.1	149.7
VHNet	9.33	58.2	58.4
LF-VSN	7.40	55.4	54.9
DRANet	16.54	130.6	112.5
本文模型	12.13	86.3	85.7

提高载密视频不可见性的同时,降低了提取秘密视频的失真程度。

3.5 消融实验

在消融实验中,本文重点关注对抗蒸馏训练方式的有效性。通过对比传统的噪声层方法与本文中使用的U-Net视频编码模拟器在两种训练方式下的表现,其中,PSNR结果如图5所示,SSIM结果如图6所示。

由图5和图6可知,本文提出的基于Unet的视频编码模拟器在训练效果上优于传统的噪声层方法。同时,对抗蒸馏训练法的结果也优于预训练视频编码器的训练方式。

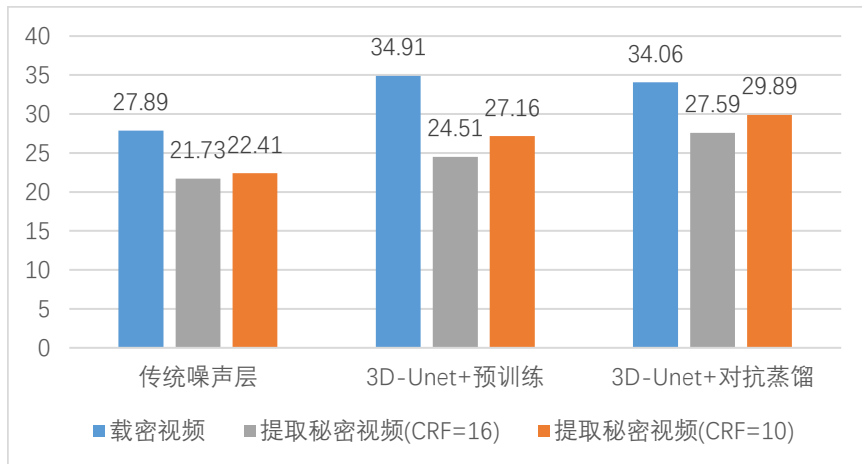


Fig. 5 Results of ablation test on the REDS dataset (PSNR)

图5 REDS数据集消融实验结果(PSNR)

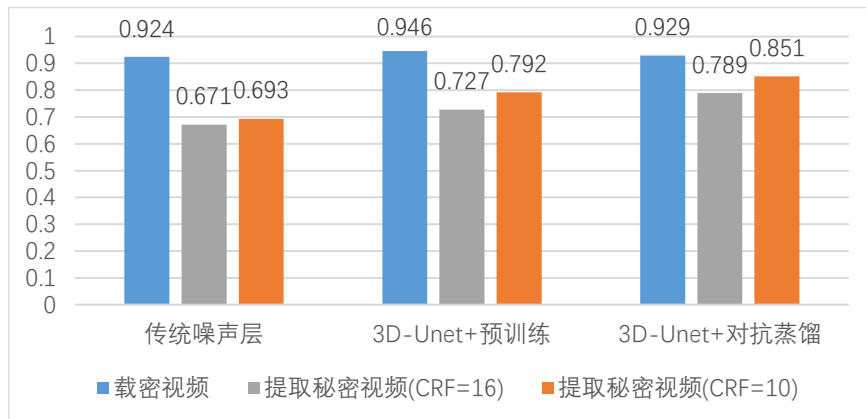


Fig. 6 Results of ablation test on the REDS dataset (SSIM)

图6 REDS数据集消融实验结果(SSIM)

实验表明,本文所提出的视频编码器和对抗蒸馏的训练方式是有效的。

图7是随着视频隐藏模型的训练层数加深,载密视频和提取秘密视频的失真程度的对比情况。随着视频隐藏模型的训练层数加深,采用传统预训练方式训练的视频编码器会使得模型性能下降。虽然训练过程中损失值呈现下降趋势,但提取的秘密视频却表现出更低的失真程度。其原因在于随着训练层数的加深,视频隐藏模型越来越倾向于拟合视频编码模拟器,却忽视了真实视频编码带来的影响。从最终结果和拟合曲线的表现来看,所提出的对抗蒸馏训练方法在视频隐藏模型的训练中展现出了显著的优势。

3.6 H.265/HEVC 编码扩展实验

H.265/HEVC 视频编码作为目前最新的视频编码标准,拥有在将视频编码比特率降低 50% 的同时保持相同程度的视频感知质量。本文选用 H.265/HEVC 视频编码作为扩展实验,用以验证本文提出的视频隐藏模型的通用性。使用 REDS 数据集生成载密视频,视频分辨率为 512×512,视频总长度为 96 帧,其中,以 8 帧作为分段嵌入秘密视频,质量因子 CRF 分别为 10 和 16,比特率 4 000 K。其结果如

表5所示。

从表5可以看出,与 H.264/AVC 视频编码相比,本文生成的载密视频在经过 H.265/HEVC 视频编码后提取的秘密视频质量有所下降。结合表1数据进行对比,在 CRF 为 10 时,PSNR 和 SSIM 指标分别下降了 1.52 dB 和 0.005;在 CRF 为 16 时,SNR 和 SSIM 指标分别下降了 1.41 dB 和 0.018,虽然性能有所下降但总体上仍处于可用范围内。实验结果表明,本文提出的抗视频编码隐写模型具备一定的视频编码扩展能力。

4 结语

本文提出了一种基于对抗蒸馏的抗视频编码的视频隐藏模型。相较于以往的视频隐藏模型,本文构建了三维可逆神经网络用于嵌入和提取秘密视频,并设计了基于三维 Unet 的视频编码模拟器来模拟视频编码器。此外,本文采用了对抗蒸馏的训练方式,以便视频编码模拟器能够更加逼真地模拟视频编码过程。该训练方式结合了生成对抗网络和知识蒸馏,在训练过程中动态地提高视频编码模拟器的模拟能力,最终提高了视频隐藏模型的抗视频编码

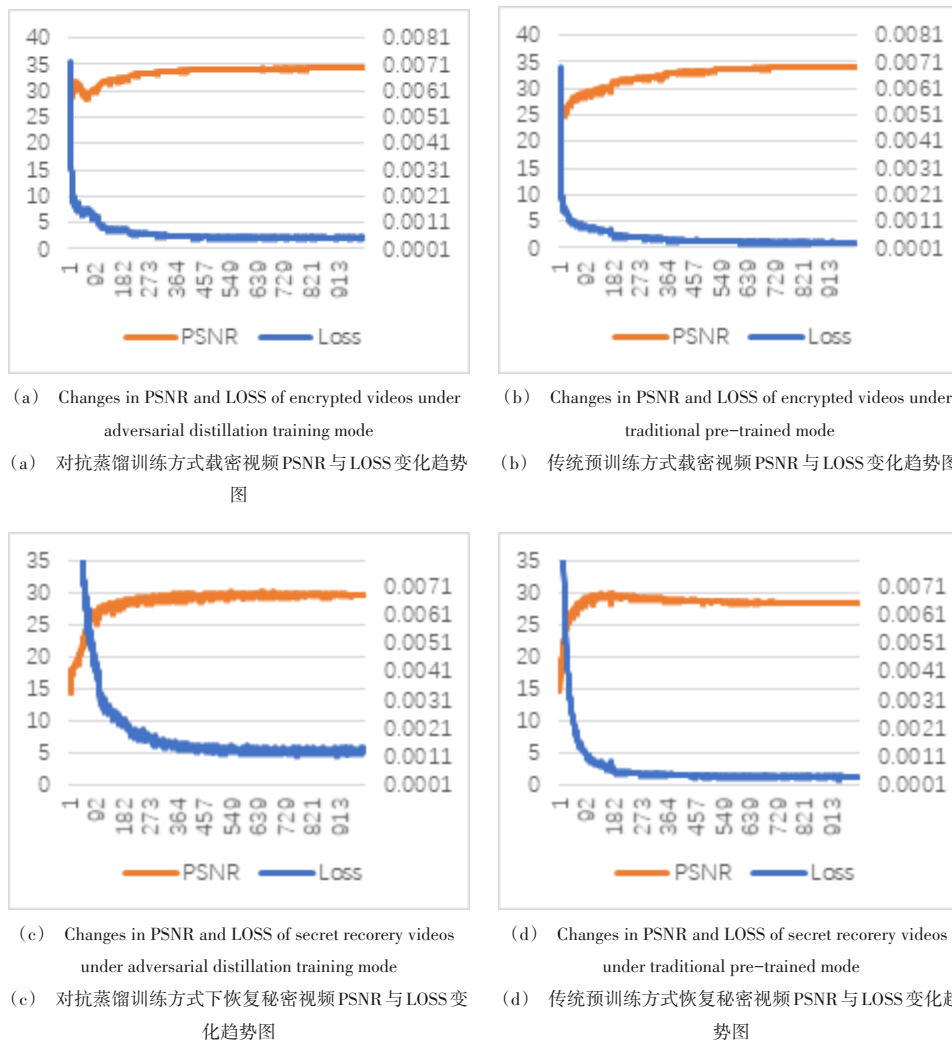


Fig. 7 The loss value and performance diagram of the container image and the reveal secret image under different training modes

图 7 不同训练方式下的载密视频和提取秘密视频的损失值与性能图

Table 5 Encoding experiment results of H.265/HEVC

表 5 H.265/HEVC 编码实验结果

CRF	载密视频		提取秘密视频	
	PSNR	SSIM	PSNR	SSIM
10	34.06	0.929	28.37	0.836
16			26.18	0.771

能力。实验数据表明,本文提出的视频隐藏模型在抗视频编码能力上优于目前的常见模型,在现实中具有更广泛的应用价值。后续工作可以从网络参数量、载密视频生成效率以及 H.265/HEVC 视频编码扩展性 3 个方面入手,进一步完善和扩展视频隐藏模型。

参考文献:

- [1] KUNHOTH J, SUBRAMANIAN N, AL-MAADEED S, et al. Video steganography: recent advances and challenges[J]. Multimedia Tools and Applications, 2023, 82(27): 41943-41985.
- [2] BOUZEGZA M, BELATRECHE A, BOURIDANE A, et al. A comprehensive review of video steganalysis[J]. IET Image Processing, 2022, 16(13): 3407-3425.
- [3] OLOYEDE A, DAUDA A, SAKA B, et al. Review and analysis on audio

steganography techniques[J]. International Journal of Engineering in Computer Science, 2024, 6(1): 22-29.

- [4] PAUL P K, CHATTERJEE R, DAS N, et al. Multimedia security: basics, its subfields, and allied areas[M]. Boca Raton: Securing the Digital World, 2025.
- [5] LIANG H, GUO L, CHEN Y B. High capacity video steganographic algorithm based on visual perception model[J]. Journal of University of Chinese Academy of Sciences, 2010, 27(1): 55-62.
梁惠, 郭立, 陈运必. 基于视觉感知模型的大容量视频隐写算法[J]. 中国科学院大学学报, 2010, 27(1): 55-62.
- [6] LUO X, LI Y, CHANG H, et al. Dvmark: a deep multiscale framework for video watermarking[J]. IEEE Transactions on Image Processing, 2025, 34: 4371-4385.
- [7] BAI X, CHEN Y, DUAN G, et al. A data hiding scheme based on the difference of image interpolation algorithms[J]. Journal of Information Security and Applications, 2022, 65: 103105.
- [8] LIU S, XU D. A robust steganography method for HEVC based on secret sharing[J]. Cognitive Systems Research, 2020, 59: 207-220.
- [9] MANDAL P C, MUKHERJEE I, PAUL G, et al. Digital image steganography: a literature survey[J]. Information Sciences, 2022, 609: 1451-1488.

- [10] WANI M A, SULTAN B. Deep learning based image steganography: a review [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2023, 13(3): e1481.
- [11] ABDULLA A A. Digital image steganography: challenges, investigation, and recommendation for the future direction [J]. Soft Computing, 2024, 28(15): 8963–8976.
- [12] MAJEED N D, AL-ASKERY A J, HASAN F S, et al. A survey on steganography and image encryption techniques [J]. Electrical Engineering Technical Journal, 2025, 2(1): 11–24.
- [13] ZHU J, KAPLAN R, JOHNSON J, et al. Hidden: hiding data with deep networks [C]//Proceedings of the European Conference on Computer Vision, 2018: 657–672.
- [14] BALUJA S. Hiding images within images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(7): 1685–1697.
- [15] WENG X, LI Y, CHI L, et al. High-capacity convolutional video steganography with temporal residual modeling [C]//Proceedings of the 2019 on International Conference on Multimedia Retrieval, 2019: 87–95.
- [16] JIA J, GAO Z, ZHU D, et al. Rivie: robust inherent video information embedding [J]. IEEE Transactions on Multimedia, 2022, 25: 7364–7377.
- [17] JAISWAL A, KUMAR S, NIGAM A. En-VStegNET: video steganography using spatio-temporal feature enhancement with 3D-CNN and Hourglass [C]//2020 International Joint Conference on Neural Networks (IJCNN), 2020: 1–8.
- [18] XU Y, MOU C, HU Y, et al. Robust invertible image steganography [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 7875–7884.
- [19] SONG B, WEI P, WU S, et al. A survey on deep-learning-based image steganography [J]. Expert Systems with Applications, 2024, 254: 124390.
- [20] KUMAR A, RANI R, SINGH S. A survey of recent advances in image steganography [J]. Security and Privacy, 2023, 6(3): e281.
- [21] LU S P, WANG R, ZHONG T, et al. Large-capacity image steganography based on invertible neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 10816–10825.
- [22] JING J, DENG X, XU M, et al. Hinet: deep image hiding by invertible network [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 4733–4742.
- [23] MOU C, XU Y, SONG J, et al. Large-capacity and flexible video steganography via invertible neural network [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 22606–22615.
- [24] ZHANG Y, NI J, SU W, et al. A novel deep video watermarking framework with enhanced robustness to h. 264/avc compression [C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023: 8095–8104.
- [25] LIU Y, XU D W, YANG L, et al. An adaptive steganography algorithm for HEVC video without intra frame distortion drift [J]. Journal of Optoelectronics·Laser, 2024, 35(3): 272.
刘勇,徐达文,杨璘,等.一种无帧内失真漂移的HEVC视频自适应隐写算法[J].光子·激光, 2024, 35(3): 272.
- [26] SHEN X, YAO H, TAN S, et al. VHNet: a video hiding network with robustness to video coding [J]. Journal of Information Security and Applications, 2023, 75: 103515.
- [27] YANG L, XU D, WANG R, et al. Adaptive HEVC video steganography based on distortion compensation optimization [J]. Journal of Information Security and Applications, 2023, 73: 103442.
- [28] SU H, LIU W X, ZHANG S Q. W-AE: a wavelet-based auto encoder for steganography [J]. Software Guide, 2023, 22(8): 178–186.
苏海,刘卫星,张淑青. W-AE: 基于小波变换的自动编码器隐写术 [J]. 软件导刊, 2023, 22(8): 178–186.
- [29] ERNAWAN F. An improved hiding information by modifying selected DWT coefficients in video steganography [J]. Multimedia Tools and Applications, 2024, 83(12): 34629–34645.
- [30] YAO Y, WANG J, CHANG Q, et al. High invisibility image steganography with wavelet transform and generative adversarial network [J]. Expert Systems with Applications, 2024, 249: 123540.
- [31] BABU J S, NIRANJANA G, RAMANA K. Adaptive sail fish optimizer based optimized region selection for robust video steganography [J]. International Journal of Information Technology, 2025, 17: 2417–2427.
- [32] CUI J, ZHENG L, YU Y, et al. Deeply-recursive attention network for video steganography [J]. CAAI Transactions on Intelligence Technology, 2023, 8(4): 1507–1523.
- [33] BISWAL M, SHAO T, ROSE K, et al. StegaNeRV: video steganography using implicit neural representation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 888–898.

(责任编辑:陈维捷)