小即是大：重新思考深度隐写的嵌入率

李涵、Hangche ng Liu、郭尚伟、周明亮、王宁、向涛和Tianwei Zhan g

arXiv:2302.11918v1 [cs.MM] 23 Feb 2023

摘要—**深度隐藏利用深度神经网络（DNNs）隐藏秘密信息，能够显著提高嵌入率并提升秘密共享的效率。现有研究主要致力于设计具有更高嵌入率或华丽功能的深度神经网络。在本文中，我们希望回答一些基本问题：如何提高深度隐藏的嵌入率，以及什么因素决定其嵌入率。为此，我们首先提出一种新颖的局部深度隐藏（LDH）方案，通过将大型秘密图像隐藏到载体图像的小型局部区域中，显著提高嵌入率。我们的方案由三个深度神经网络组成：隐藏网络、定位网络和提取网络。我们使用隐藏网络将秘密图像转换为小型、不可感知的紧凑秘密代码，并将其嵌入载体图像中的随机局部区域。定位网络通过识别隐写图像中秘密代码的位置来辅助提取过程，而提取网络则从这些已识别的局部区域中恢复所有完整尺寸的秘密图像。我们的LDH实现了极高的嵌入率，即 16×24 每像素比特数，并表现出对常见图像失真的出色鲁棒性。我们还开展了全面实验，在各种系统设置下评估我们的方案。此外，我们利用不同的图像复原算法，对嵌入率与图像质量之间的权衡进行了定量分析。**

索引词—**深度隐藏、嵌入率、深度神经网络、定量分析**

一、引言

数ATA据隐藏或隐写术是一种将秘密消息隐藏在看似无害的载体对象（如图像、视频或其他计算机文件）中的艺术 [1], [2], [3], [4], [5],[6]。这项技术使人们能够在公共信道中进行秘密共享，而不会引起他人对隐藏信息传输的怀疑。数据隐藏方案通常包括两个阶段：隐藏和提取。在隐藏阶段，秘密消息的所有者将消息嵌入载体媒介（例如图像）中，生成隐写媒介，同时保持载体媒介的保真度。在提取阶段，接收方从隐写媒介中提取嵌入的秘密消息。本文重点讨论图像中秘密消息的隐藏与提取。

在评估数据隐藏方案时，嵌入率是衡量秘密共享效率的最重要指标之一，可用每像素比特数（bpp）量化。人们为提高嵌入率做出了许多努力。例如，HUGO [7] 通过设计编码算法，最小化经过适当定义的失真，从而提高嵌入率。遗憾的是，此类传统数据隐藏方案 [7], [8], [9], [10], [11], [12], [13], [14], [15] 通常嵌入率较低，例如，HUGO 的嵌入率仅约为 0.5 每像素比特数。嵌入率如此之低，意味着 HUGO 只能用于传输简短的文本消息。据我们所知，这也是大多数传统数据隐藏方案 [8], [9], [13], [14], [15],普遍存在的问题，尤其是基于频域的方案 [10], [11], [12]。

近年来，深度学习技术已应用于数据隐藏（即深度隐藏），通过深度神经网络（DNN）隐藏和提取秘密消息，并能显著提高嵌入率。Baluja等人 [16] 首次利用深度神经网络，通过将一幅全尺寸图像隐藏在另一幅图像中，实现了高达 24 每像素比特数的嵌入率。此后，研究者提出了更多深度隐藏方案 [17], [18], [19], [20] ，以达到相近或更高的嵌入率。例如，Zhang等人[17]提出了一种深度隐藏方案，通过训练三对隐藏与揭示网络，将三幅秘密图像隐藏在一幅载体图像中（即 3×24 每像素比特数）。

现有深度隐藏方案已实现较高的嵌入率，既能提高秘密图像的图像质量 [18], [19], 也能提供将超链接隐藏在图像中 [21], 、光场消息传递[22]等新颖功能。然而，对于这一现象的成因及影响嵌入率的因素，鲜有分析。本文旨在回答两个更基本的问题：1）为什么深度神经网络能够大幅提高嵌入率？2）嵌入率还能否进一步大幅提高？以及 3）嵌入率与图像质量之间是否存在权衡？。

对于第一个问题，我们发现，深度神经网络经过训练，可将秘密图像编码为不可感知代码，再将其嵌入载体图像中。这种高效的编码技术能够有效缩小秘密图像的规模，并显著提高相应隐藏方案的嵌入率。基于这些见解，我们通过设计一种新颖的局部深度隐藏（LDH）来解决第二个问题，并将嵌入率进一步提升至惊人的水平，16×24 每像素比特数。具体而言，我们的 LDH 使用经过精心设计的深度神经网络，将一幅全尺寸秘密图像编码为较小的不可感知紧凑秘密代码，这是提高嵌入率的关键。该小型不可感知紧凑秘密代码会随机嵌入载体图像的一个局部区域。由于载体图像中其余像素可用于隐藏其他秘密图像，我们的 LDH 能够嵌入比现有深度隐藏方案更多的图像；现有方案会将一幅全尺寸秘密图像嵌入载体图像的所有像素中。

我们的 LDH 由三个深度神经网络组成：隐藏网络、定位网络和提取网络。隐藏网络将给定的秘密

李涵和刘航成贡献相同。李涵、刘航成、郭尚伟、MingliangZhou、NingWang和向涛来自中国重庆大学计算机学院。张天威来自新加坡南洋理工大学计算机科学与工程学院。

图像转换为一种微小且不可感知的紧凑秘密代码（看起来像一张尺寸较小的噪声图像），然后将其添加到载体图像的随机局部区域中，从而生成相应的隐写图像。定位网络会识别隐写图像中所有嵌入的秘密代码的位置，并输出一张二值位置图；借助该图，提取网络可以从隐写图像中确定的局部区域恢复出所有完整尺寸的秘密图像。我们的 LDH 还可以通过同一组隐藏与揭示网络完成所有秘密图像的嵌入与恢复，无需训练多组隐藏与揭示网络。实验结果表明，我们的 LDH 在保持载体图像和安全图像质量的同时，能够实现高得多的嵌入率。

为回答最后一个问题，我们构建了一个秘密共享实验场景，并从消息接收方的角度进行了深入分析。具体而言，我们假设接收方能够使用不同的算法来提升揭示出的秘密图像的质量。我们测量了各种深度隐藏方案在图像质量达到特定阈值时所能实现的最高嵌入率。我们发现，深度隐藏方案的嵌入率会随质量阈值的不同而变化。随着嵌入率的提高，现有隐藏方案所揭示结果的质量会逐渐下降，而我们的 LDH 则不会出现这一现象。此外，如果接收方拥有强大的图像复原模型来提升秘密图像的质量，深度隐藏方案的嵌入率也可能随之提高。

我们将主要贡献总结如下：

- 我们提出了一种新颖的局部深度隐藏（LDH）方案，可将秘密图像编码为小巧且难以察觉的秘密代码，从而显著提高嵌入率。
- 我们提出了一种两阶段训练策略，以提升 LDH 的性能。
- 我们进行了广泛的实验，以评估我们的 LDH 的优越性。
- 我们进行了深入研究，以定量分析嵌入率与图像质量之间的权衡。

本文其余部分的组织结构如下：第二节介绍数据隐藏及相关工作。第三节详细说明我们提出的方案。第四节在各种系统设置下评估我们的方案，随后对嵌入率与图像质量之间的权衡进行定量分析。第六节对本文进行总结。

II. 背景与相关工作

为便于描述数据隐藏方案，我们分别使用 S 和 C 表示秘密消息和载体图像。ϵ_1 和 ϵ_2 是两个超参数。D 用于衡量两个给定输入之间的差异。数据隐藏方案由两个必需的算法组成：

- **隐藏算法** (H)将 S 嵌入 C 中，且不影响 C的视觉质量，并生成隐写图像 C'（即 $C' = H(C, S)$），其具有 $D(C', C) \leq \epsilon_1$；
- **提取算法** (R)从 C' 中提取嵌入的秘密消息，并将其作为 $S' = R(C')$，且需要满足$D(S', S) \leq \epsilon_2$。

数据隐藏方案的性能通常通过一个载体像素能够承载的比特数来衡量，即每像素比特数（每像素比特数）。每像素比特数的数值越高，表示相应的数据隐藏方案具有更高的消息容量（即更高的嵌入率）。

A. 低嵌入率的数据隐藏

一些数据隐藏方案将秘密消息嵌入频域，例如离散余弦变换（DCT）域 [23], [24], [25], 离散小波变换（DWT）域 [26], [27], 离散傅里叶变换（DFT）域 [12], [28] 以及整数小波变换（IWT）域[29], [30]。然而，这类频域数据隐藏方案只能达到较低的嵌入率，例如每像素比特数为 2 或更低。例如，Attaby 等人 [25]提出了 DCT-M3算法，该算法利用两个 DCT 系数之差对 3 取模来嵌入两位秘密消息。Atawneh 等人 [27] 采用菱形编码方法实现了一种基于 DWT 的隐藏方案，从而提升了隐写图像的质量。McKeon 和 Robert [12] 提出了一种基于 DFT 的方案，为在影片中隐藏数据提供了一种精巧的方法。Kalita 等人 [30] 开发了一种基于 IWT 的方法，仅利用高频子带隐藏秘密消息，以抵御隐写分析。尽管基于频域的方案通常对常见干扰操作（如压缩）具有良好的鲁棒性，但其计算复杂度很高，且只能实现有限的嵌入率。

相反，其他数据隐藏方案将秘密消息嵌入空域，例如最低有效位（LSB）[8], [31], [32], 、像素值差分（PVD） [33], [34],、直方图平移 [35], [36], [37], 、基于扩展的方法 [38], [39],和多位平面 [40], [41]。与基于频域的方法相比，这些方法的嵌入率更高，但通常鲁棒性较差，也更容易受到隐写分析的影响 [42],[43], [44], [45], [46]。空域方法直接对像素进行操作，因此通常具有更高的嵌入率（例如超过 2 bpp）。近期研究也采用深度学习技术来提高嵌入率。例如，Hayes 等人 [47]设计了一种基于生成对抗网络的数据隐藏方法，以提升隐写图像的质量。Ghamizi 等人 [48] 利用深度神经网络和对抗攻击将秘密信息嵌入图像，并在一定程度上提高了嵌入率。Hu 等人 [49] 提出了一种无嵌入隐写术（SWE）方法，该方法利用训练好的生成器神经网络模型，根据噪声向量生成隐写图像，从而进一步增强抵御检测算法的能力。Ray 等人 [50] 使用基于深度监督的边缘检测器，将更多比特嵌入边缘像素中。Tancik 等人 [21] 设计了一种名为 StegaStamp 的新框架，以在物理世界中实现消息的隐藏与提取。然而，这些基于深度学习的数据隐藏方案侧重于将二进制消息嵌入载体图像，且嵌入率仍然较低。

B. 高嵌入率数据隐藏

更高的嵌入率意味着可以通过同一幅载体图像传递更多秘密消息，从而提高秘密共享的效率。现有的高嵌入率数据隐藏方案使用深度神经网络将全尺寸图像隐藏在图像中，嵌入率可超过 24 bpp。根据模型结构的不同，我们可以将这些方案分为三类：编码器-解码器结构 [16], [51], [52], [19], [17], [53]、生成对抗网络（GAN）结构 [54], [55], [18], 和可逆神经网络（INN）结构 [20], [56]。

编码器-解码器结构。 由于数据隐藏与揭示过程分别与编码和解码过程完全对应，因此编码器-解码器结构适用于数据隐藏模型。例如，Baluja等人 [16] 首次设计了一种将图像隐藏在图像中的框架，该框架使用预处理网络，通过隐藏网络将秘密图像嵌入载体图像。Zhang等人 [17] 移除了预处理网络，并设计了一种大容量数据隐藏框架，其中秘密图像的编码过程独立于载体图像的选择，且秘密代码可以隐藏在任何载体图像中。他们随后训练了三对隐藏与揭示网络，将嵌入率提高到72每像素比特数。类似地，Das等人 [53] 提出了MISDNN，该方法采用若干对准备网络、隐藏网络和揭示网络来实现多图像隐藏。

基于GAN的结构。 在这类深度隐藏中，隐藏和揭示过程都可以由专门的生成器实现，而判别器则负责使生成结果尽可能接近原始结果。Meng等人 [55] 采用 CycleGAN [57] 实现图像间的图像隐藏，并通过对抗训练提高数据隐藏的安全性。类似地，Yu等人 [18]也在深度隐藏中采用了GAN，并进一步引入注意力机制以提升图像质量。

INN结构。这类深度隐藏方案利用 INN [58], [59] ，并在隐藏与揭示网络之间共享参数，而不是训练两个独立的网络。Lu等人 [20] 使用 INN 同时实现隐藏与揭示网络的功能，并通过沿通道维度拼接多幅秘密图像来实现多图像隐藏。Guan等人 [56]开发了一种可逆隐藏神经网络（IHNN），可以多次级联。隐藏多幅秘密图像时，IHNN 能够根据先前的图像隐藏结果来指导当前图像的隐藏。

尽管现有深度隐藏方案显著提升了隐藏性能，尤其是嵌入率，但仍无法充分回答基本问题。为了进一步分析其功能并提高嵌入率，我们将在下文提出一种新颖的局部深度隐藏方案。

III. 方法论

洞见。 深度隐藏的高嵌入率令人惊讶，同时我们也在思考嵌入率得以提升的原因，以及如何进一步提升嵌入率。通过比较传统数据隐藏与深度隐藏，我们发现，深度神经网络的编码能力对于提升嵌入率至关重要，因为图像具有很高的冗余度，而且深度神经网络能够自动将秘密图像的信息嵌入载体图像中，同时不影响其保真度。

为了展示这一现象，我们设计了一种局部嵌入策略来解释深度隐藏的工作原理，同时在不影响载体图像视觉质量的前提下，实现更高的嵌入率。具体而言，我们提出一种新颖的局部深度隐藏（LDH）方法：利用隐藏网络将秘密图像转换为小型、不可感知的紧凑编码（例如，编码大小仅为秘密图像大小的$_{1\ 16}$）。我们要求提取网络恢复出对应的完整尺寸、完整内容的秘密图像。显然，这种小型、不可感知的紧凑编码承载了秘密图像的全部信息，且能够被提取网络识别；由于编码能力有限，传统数据隐藏难以做到这一点。此外，随着秘密代码变得更小，我们可以通过精心设计损失函数和训练过程，在载体图像的不同局部区域中隐藏更多图像，从而提高嵌入率。

流程。我们在图1中展示了所提出的 LDH 的流程，该流程由两个模块组成：隐藏模块和揭示模块。在隐藏模块中，我们设计了一个隐藏网络（H），将给定的秘密图像（S）转换为一个尺寸小于 S 的小型不可感知紧凑编码（S_e）。随后，我们将紧凑秘密代码 S_e 隐藏在 C 的随机局部区域中，生成对应的隐写图像（C'）。请注意，我们的方法具有通用性，同一个 S_e 可以隐藏在不同的载体图像中。在揭示模块中，包含两个深度神经网络：定位与揭示网络。我们首先设计一个定位网络 P，通过生成二值图（B'）来确定 S_e 在 C'中的位置。接着，根据 B' 裁剪 C' 中确定的局部区域，并将裁剪结果输入揭示网络（R），以恢复对应的完整尺寸秘密图像（S'）。

*A. LDH*网络

隐藏网络。现有的深度隐藏方案 [16], [17],[18]通常会将一幅或多幅秘密图像隐藏在一幅尺寸相同的载体图像中。然而，由于多幅秘密图像的信息被嵌入载体图像的所有像素中，恢复出的秘密图像的鲁棒性和视觉质量都会受到影响。因此，很难以这种全局方式提高深度隐藏的嵌入率。此外，人们无法清楚地了解深度隐藏的工作原理。为了解决这些问题，我们考虑采用局部嵌入，并设计了一种新颖的隐藏网络，将完整尺寸的秘密图像转换为小型、不可感知的紧凑编码，如图1所示。

在隐藏网络中，我们首先采用一些下采样层来缩小 S 的尺寸，以生成 S的中间表示。随后，这些中间表示会由 H 进一步编码，输出最终的小型、不可感知的紧凑秘密代码 S_e。本文使用缩放因子 ω 来控制 S_e 的尺寸，其中 ω 为

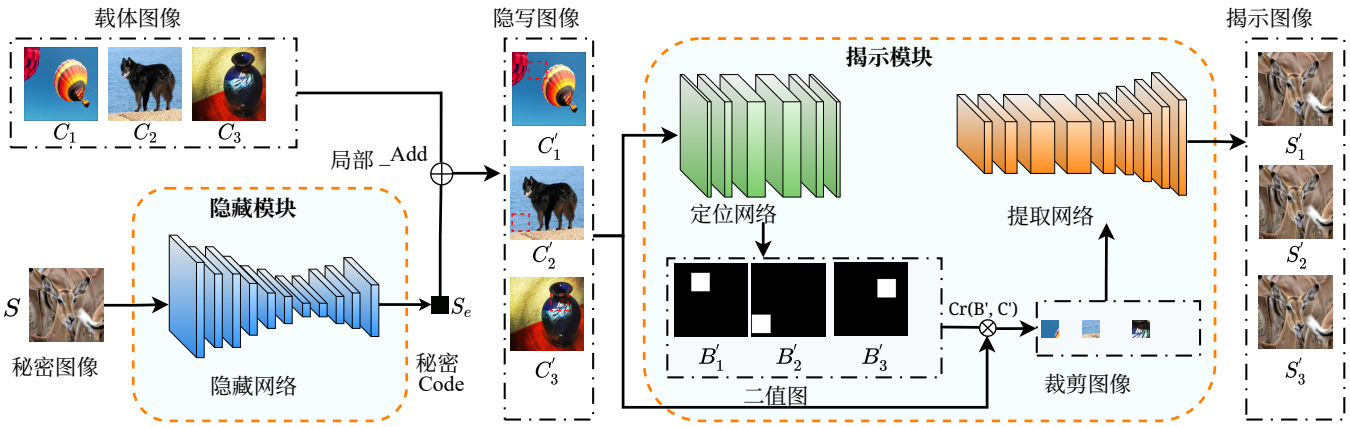


图1. 提出的 LDH 框架。

S_e 的宽度与 S的宽度之比。得到S_e后，如图1所示，我们通过逐元素相加将其隐藏在给定载体图像 C 的随机局部区域中。完整的隐藏过程可形式化表示为

$$C' = Local_Add(C, H(S, \omega)), \tag{1}$$

其中，$Local_Add$ 表示加法运算。请注意， ω 决定了嵌入率的上限。

定位网络。 由于微小且不可感知的代码S_e 被局部、随机地嵌入到 C 中，我们需要确定 S_e 在 C'中的位置。为此，我们设计了定位网络 P ，以自动识别嵌入了 S_e的区域，从而帮助提取网络重建秘密图像。定位网络P 以 C' 为输入，并通过生成二值图 B'来确定 S_e 在C' 中的位置。具体而言，我们在 B' 中用 1 表示 C' 的对应像素属于嵌入区域，用 0 表示相反情况。整个定位过程为

$$B' = P(C'). \tag{2}$$

提取网络。 在以往的工作中， [16], [19], [17],[18], 提取网络会从与 S'尺寸相同的 S_e 中恢复 S' 。尽管这些工作实现了 24 bpp 或更高的嵌入率，但仍未充分发挥提取网络的潜力。在 LDH 中，我们充分利用提取网络的能力，使其能够在参数中存储除 S_e 之外的额外图像信息。因此，我们的提取网络能够从小巧紧凑的S_e中恢复完整尺寸的 S' 。具体而言，在 R中，我们使用底层对S_e 进行预处理，以获得相应的中间表示，同时保持尺寸不变。然后，我们使用R 的顶层进一步转换该表示，并逐步增大其尺寸，直到达到 S的原始尺寸。提取过程概述如下

$$S' = R(Cr(C', B')), \tag{3}$$

其中， $Cr(C',\ B')$ 表示在 B' 的指导下对 C' 执行裁剪操作。

B. 训练细节

我们对这三个网络进行联合训练，以确保秘密共享的不可感知性和安全性。具体而言，我们有三个目标：1）最大限度地减少隐藏操作对载体图像的负面影响；2）准确识别嵌入区域的位置；3）最大限度地减小原始秘密图像与揭示出的秘密图像之间的差异。接下来，我们将介绍用于实现这些目标的三个损失函数。

隐藏损失。 我们使用隐藏损失（L_H）来约束H，从而确保 S_e的不可感知性。使 C' 接近 C 的一种常见方法是最小化 L_p 距离，即C' 与 C [16]之间的距离， [17]。因此，在我们的实验中，我们将L_H 设为

$$L_H = \|C - C'\|_p. \tag{4}$$

定位损失。 准确识别嵌入区域的位置对我们的方案至关重要，因为这会直接影响最终的揭示结果。在训练 P 之前，我们将所有嵌入秘密代码S_e 的真实位置记录在二值图 B （即真实标注）中，其中 B_{ij} 表示 C' 中对应的像素携带秘密信息，否则为 0。在训练过程中，我们优化 P ，以减小真实标注与 P 的预测图 B' 之间的定位损失（L_P），即

$$L_P = \|B - B'\|_p. \tag{5}$$

揭示损失。 R 负责从裁剪后的隐写图像 C'中揭示 S' ，其结果应尽可能接近原始秘密图像 S。因此，我们还以 L_p-范数的形式设计了相应的揭示损失（L_R），即

$$L_R = \|S - S'\|_p. \tag{6}$$

结合上述三个损失项，我们可以得到用于训练 LDH 中三个深度神经网络的最终损失函数，即:

$$L_T = \lambda_1 L_H + \lambda_2 L_P + \lambda_3 L_R, \tag{7}$$

其中， λ_1、 λ_2、 λ_3 是三个超参数。

训练策略。 为了更好地训练深度神经网络，我们采用两阶段训练策略。具体而言，我们将整个训练过程分为预训练阶段和协同训练阶段。在预训练阶段，我们设置 $\lambda_2 = 0$。这是因为，在

训练初期，P 无法正确确定嵌入区域。我们不要求 R 从非嵌入区域中恢复 S'，否则会导致训练停滞。因此，在预训练阶段，我们不考虑 P。请注意，为了训练 H 和 R，我们在此阶段使用 B（位置的真实标注）对 C' 进行裁剪。之后，我们开始协同训练 H、P和 R。鉴于 H 和 R 已得到良好训练，我们在协同训练阶段主要着重训练 P。因此，我们将 λ_2 设为较大值，而在协同训练中将 λ_1 和λ_2 设为较小值。

C. 多图像隐藏

得益于局部嵌入策略，LDH 能够实现多图像隐藏，从而提高嵌入率。如图 2 所示，四张秘密图像由 H 编码为相应的不可感知紧凑编码，然后嵌入到 C 的不同局部区域中。值得注意的是，我们的 LDH 也支持将多张图像隐藏在同一局部区域或重叠的局部区域中；这只需在训练过程中将此类隐藏样本纳入真实标注即可。为简便起见，我们仅考虑将多张图像嵌入载体图像的不同区域。我们的定位网络 P 能够识别所有嵌入的秘密图像的位置，并输出包含多个白色区域的二值图。最后，R 从 C' 的相应裁剪区域中恢复所有完整尺寸的秘密图像。

理论上，如果通信双方事先商定 ω^2 的取值，*LDH* 最多可同时支持 $\omega^2 \times 24$ 张秘密图像（即每像素比特数）。与现有的多图像隐藏方案（例如 UDH [17]）不同，LDH 只需使用一套 H、R 和 P，即可隐藏多张图像，而无需训练多个网络对。因此，我们的 LDH 效率更高。LDH 中的多图像隐藏可表示为

$$C' = Local_Add\left(C, \{H(S^{(k)}, \omega)\}\right), \tag{8}$$

其中 $S^{(k)}$ 是第 k张秘密图像。相应的揭示过程为

$$\{S'^{(k)}\} = R(Cr(C', P(C'))). \tag{9}$$

除了显著提高嵌入率之外，LDH 还有另一项优势：LDH 可以通过将 S_e 嵌入到 C中纹理丰富的局部区域，进一步降低他人检测到共享秘密存在的风险。例如，可以使用恰可察觉差异[60] (JND)技术来衡量人眼对不同区域失真的敏感度，并找出最适合隐藏秘密图像的位置，从而获得

C'，具有更好的视觉质量。

IV. 实验

A. 配置

数据集。我们使用 ImageNet 和 COCO 评估所有涉及的隐藏方案 [16], [17], [18], [20], [56]。对于每个数据集，我们随机选取 10,000 张图像，其中 8,000 张

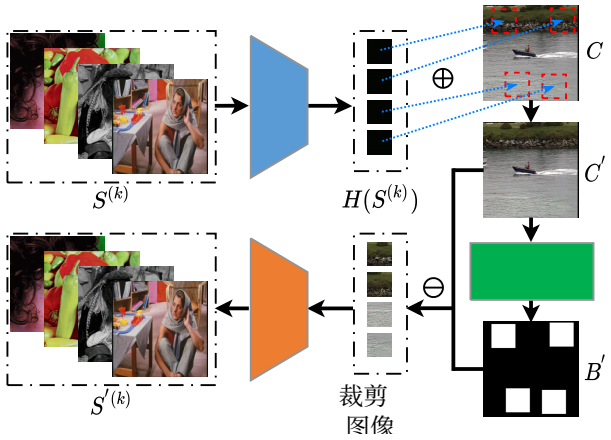


图2. 使用我们的 LDH 将四个不同的 S 隐藏到一个 C 的不同位置中。所有秘密图像都能够以高质量被检测和恢复。

用于训练，1,000 张用于验证集，1,000 张用于测试集。所有选取的图像均调整为 1024×1024，且像素值归一化至 $[0, 1]$ 范围内。

LDH 和基线方法的实现。在 LDH 的实现中，我们首先考虑 $\omega = 4$（即，由 H输出的 S_e大小为 256×256），并在后续小节中讨论ω 的其他取值。对于 H，我们采用一个卷积层和一个最大池化层构建其浅层，以获取小尺寸特征图。随后，我们采用 UNet [61] 构建其剩余层，用于生成尺寸较小且不可感知的秘密代码，即紧凑编码 S_e。为生成二值图，我们将网络 P 设计为具有六个卷积层的全卷积网络，使其输入和输出的大小保持一致。至于网络 R，我们同样使用六个卷积层构建其浅层，以对输入进行预处理，并采用 [62]中提出的网络逐步增大特征图的尺寸，直到达到原始秘密图像的大小。如前一节所述，在预训练阶段，我们设置 $\lambda_1 = 0.25$、$\lambda_2 = 0$和 $\lambda_3 = 0.75$，以获得训练良好的 H 和 R。在协同训练阶段，主要目标是训练能够准确定位 S_e位置的 P，因此我们设置$\lambda_1 = 0.1$、$\lambda_2 = 0.8$ 和 $\lambda_3 = 0.1$。此外，我们将初始学习率设为 0.001，并在训练过程中每经过 30 个训练轮次便将其缩小为原来的 0.1 倍。所有网络均使用 Adam优化器 [63]进行训练。

我们还考虑了一些基线方法进行比较，包括 DDH [16], UDH [17], HCVS [51] 和 MISDNN [53]。我们分别按照公开可用的实现[1,2,3,4] 来实现这四种基线方法。对于 DDH [16], UDH [17],和 HCVS [51]，我们还通过在通道维度上拼接多张秘密图像，实现了多图像隐藏。对于 MISDNN [53],，我们通过增加或减少网络对的数量来调整嵌入率。

评估指标。为量化所有相关方案的性能，我们使用三个视觉指标来衡量 C 与 C'（或 S 与 S'）之间的距离，包括平均

[1]https://github.com/arnoweng/PyTorch-Deep-Image-Steganography [2] https://github.com/ChaoningZhang/Universal-Deep-Hiding [3] https://github.com/muziyongshixin/pytorch-Deep-Steganography [4] https://github.com/m607stars/MultiImageSteganography

表 I 性能比较。↑表示数值越高越好，反之亦然。

方法	封面图/隐写图对						秘密图/恢复图对					
	ImageNet			COCO			ImageNet			COCO		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
DDH [16]	2.552	0.961	37.963	2.417	0.978	37.942	2.261	0.991	39.055	1.956	0.989	39.614
UDH [17]	2.031	0.979	40.035	2.121	0.981	39.746	2.850	0.969	35.796	2.694	0.975	36.708
MISDNN [53]	2.607	0.974	37.413	2.411	0.980	37.780	4.309	0.970	33.065	4.301	0.969	33.146
HCVS [51]	2.566	0.955	37.571	2.531	0.961	37.763	2.243	0.989	39.032	2.253	0.982	38.978
Ours	0.346	0.994	43.887	0.292	0.996	45.118	3.809	0.951	33.169	3.974	0.947	32.963

方法	载体/隐写图像 2 对						秘密/恢复图像 2 对					
	ImageNet			COCO			ImageNet			COCO		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
DDH [16]	3.165	0.964	36.150	3.179	0.964	35.815	3.190	0.984	35.960	3.128	0.982	35.812
UDH [17]	3.398	0.954	35.940	3.376	0.953	36.006	4.632	0.953	32.435	4.502	0.956	32.846
MISDNN [53]	6.479	0.935	30.086	6.040	0.936	30.177	7.035	0.934	29.089	7.046	0.931	28.940
HCVS. [51]	3.688	0.953	34.839	3.808	0.953	34.331	3.467	0.977	35.412	3.517	0.968	35.092
Ours	0.657	0.987	41.188	0.572	0.989	42.185	3.809	0.951	33.169	3.974	0.947	32.963

方法	载体/隐写图像 3 对						秘密/恢复图像 3 对					
	ImageNet			COCO			ImageNet			COCO		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
DDH [16]	4.160	0.940	33.824	4.270	0.937	33.641	3.806	0.978	34.391	3.846	0.978	34.340
UDH [17]	4.120	0.933	34.392	4.145	0.935	34.336	6.430	0.936	30.321	6.110	0.932	30.375
MISDNN [53]	11.102	0.886	25.694	9.647	0.893	26.407	10.708	0.892	25.682	10.17	0.903	25.957
HCVS [51]	5.080	0.912	32.301	5.056	0.930	32.177	3.802	0.971	34.616	3.871	0.965	34.130
Ours	0.983	0.982	39.572	0.988	0.984	39.651	3.809	0.951	33.169	3.974	0.947	32.963

方法	载体/隐写图像 4 对						秘密/恢复图像 4 对					
	ImageNet			COCO			ImageNet			COCO		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
DDH [16]	5.758	0.927	31.118	5.611	0.923	31.147	6.679	0.964	29.909	6.440	0.962	30.061
UDH [17]	4.740	0.915	33.163	4.755	0.917	33.139	10.785	0.891	25.725	10.414	0.882	25.767
MISDNN [53]	16.38	0.823	21.797	18.38	0.817	20.862	13.667	0.841	23.243	12.117	0.855	24.095
HCVS [51]	5.076	0.919	31.948	5.331	0.921	31.803	6.806	0.947	29.505	6.817	0.955	29.990
Ours	1.377	0.974	37.901	1.140	0.983	39.283	3.809	0.951	33.169	3.974	0.947	32.963

像素差异（平均像素差）、峰值信噪比（PSNR）和结构相似性指数（SSIM）。

- **平均像素差**。平均像素差是图像之间的平均 L_1-范数距离，可表示为

$$APD = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}|I_{i,j} - K_{i,j}|, \tag{10}$$

其中，I 和 K 分别是参考图像（C 或 S）和对应的失真图像（C' 或 S'）。H和 W 分别是图像的高度和宽度。坐标（i，j）表示像素 $I_{i,\ j}$（或 $K_{i,\ j}$）的位置。

- **峰值信噪比**。峰值信噪比是一种常用指标，用于量化信号重建的质量，在相关研究中得到广泛应用。它基于 L_2-范数构造，表示为

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right), \tag{11}$$

其中

$$MSE = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}[I_{i,j} - K_{i,j}]^2, \tag{12}$$

且 MAX_I 表示图像的最大可能像素值，在实验中为 1。

- **结构相似性指数**。结构相似性指数主要根据感知到的结构变化来衡量两幅图像之间的相似性信息。它也常用于相关研究中，可表示为

$$SSIM = \frac{(2\mu_I\mu_K + c_1)(2\sigma_{IK} + c_2)}{(\mu_I^2 + \mu_K^2 + c_1)(\sigma_I^2 + \sigma_K^2 + c_2)} \tag{13}$$

其中，μ_I（或 μ_K）是 I（或 K）的均值，σ_I（或σ_K）是 I（或 K）的方差，σ_{IK} 是 I 与 K之间的协方差。c_1 和 c_2 用于保持稳定性，其中 $c_1 =$（k_1L）^{2}，且 $c_2 =$（k_2L）^{2}。L 与 MAX_I含义相同。默认情况下，我们设定 $k_1 = 0.01$和$k_2 = 0.03$。APD 值越小，或 PSNR 和 SSIM 值越大，生成图像的质量就越好，这意味着更好的隐藏或揭示性能。

B. 总体评估

定量结果。我们评估了所有相关方案在载体图像中隐藏 1、2、3 和 4 张秘密图像的效果，并在表 I 中展示了数值结果。从表 I 可以观察到，在隐藏阶段，我们的 LDH 性能明显优于基线方法，其生成的 C' 在三个指标上均表现最佳。借助紧凑秘密代码，即使在载体图像中隐藏多达 4 张全尺寸秘密图像后，LDH 仍能将 C' 的质量保持在很高的水平（PNSR >37 dB）。

在提取阶段，LDH 同样表现出优势。从表 I 可以看出，LDH 中 S' 的质量不受秘密图像数量的影响。相反，

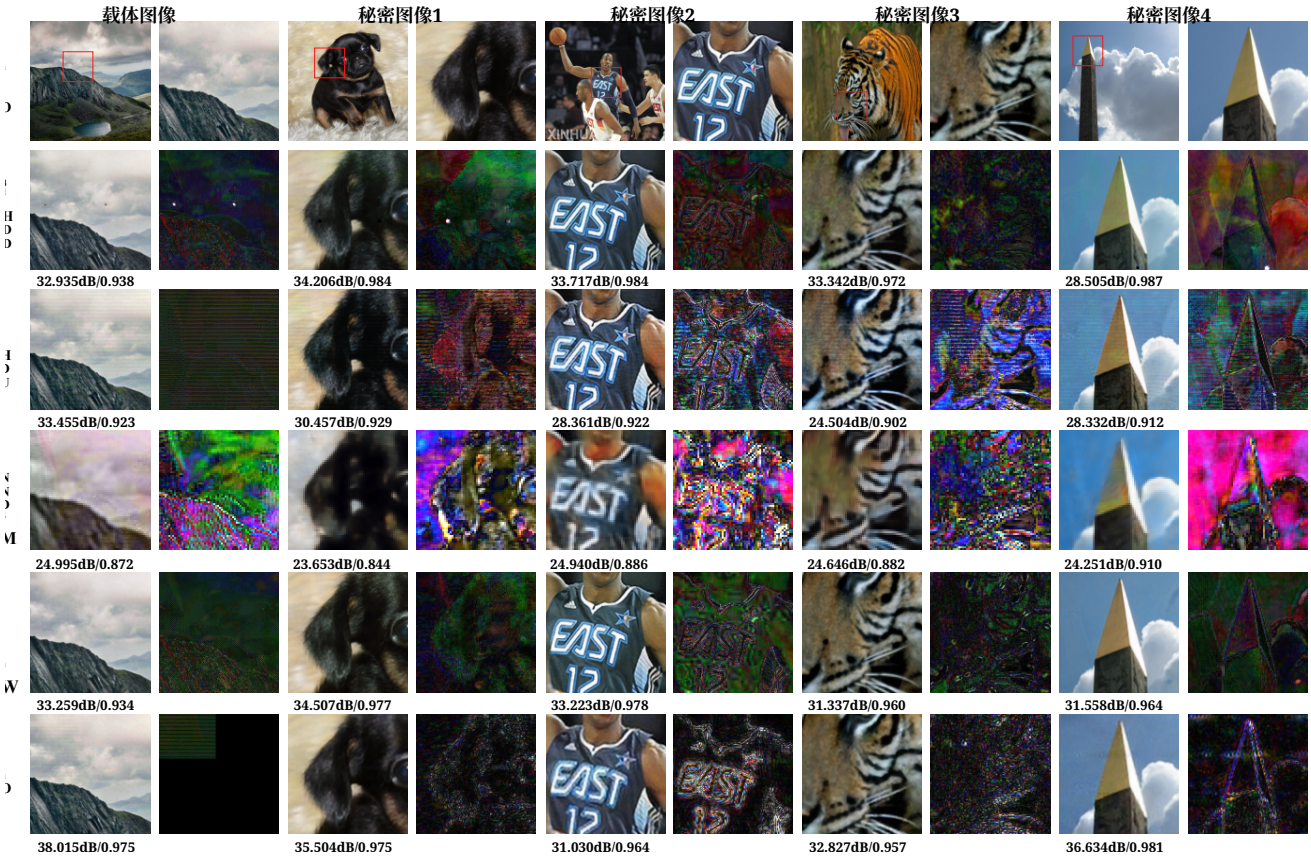


图3. 将四张秘密图像隐藏在一张载体图像中的可视化示例。顶行前两张图分别是原始载体图像，以及红框区域的对应放大图。顶行其余图像是揭示图像及其局部区域放大图。从第二行到最底行，奇数列中的图像是标记的隐藏方案生成的隐写图像或揭示的秘密图像的局部区域放大图，偶数列中的图像是隐写图像与载体图像（或揭示的秘密图像与原始秘密图像）之间的差异图像。为便于观察，图中每张差异图像的像素值均放大了 7 倍。

这些基线方法不适合嵌入更多图像，因为随着嵌入的秘密图像数量增加，秘密图像的质量会下降。LDH 与基线方法之间存在这一差距的主要原因是，LDH 中每幅秘密图像的隐藏和提取过程彼此独立。因此，提取不同的秘密图像时不会相互影响，而基线方法的多图像隐藏则会导致通道之间相互影响。在仅隐藏一幅秘密图像的情况下，LDH 的提取质量略低于 DDH [16],UDH [17] 和 HCVS [51],，但与 MISDNN [53]的提取质量相近。我们认为，这是因为与完整尺寸、不可感知的秘密代码相比，紧凑秘密代码会造成部分秘密信息丢失。

定性结果。 我们展示了图3中的一些可视化结果，以进一步体现 LDH 的优越性。首先，在隐藏阶段，LDH 对载体图像造成的扰动最小（见图3第2列）。基线方法采用全局隐藏，因此隐写图像平滑区域中的扰动很容易被察觉，例如天空中的扰动。具体而言，DDH [16]可能会在局部区域产生噪点；UDH [17]会生成带有不理想水平条纹的隐写图像；MISDNN [53]会造成严重的颜色失真；HCVS [51] 则会在秘密图像边缘产生明显的失真。尽管 LDH 仍会产生水平条纹，但我们可以选择纹理丰富的局部区域进行隐藏，避免这些条纹轻易被发现。例如，在图4中，我们使用 JND方法 [60]来确定更合适的局部区域（如树林）以承载秘密信息，从主观和客观角度来看，这都能提升隐写图像的质量。

在提取阶段，LDH 仍表现出令人满意的效果。如图3所示，我们将四张不同的秘密图像隐藏在一张载体图像中。从揭示的秘密图像可以看出，LDH 实现了最高的视觉质量。相比之下，其他方案揭示的秘密图像中出现了伪影和色彩失真。随着秘密图像数量增加，这些不理想的失真会变得更加严重（见图5）。幸运的是，由于局部嵌入的特性，我们的 LDH 不受这一问题的影响。

我们还在图 5 中展示了 DDH [16], UDH [17] 和我们的 LDH 在不同嵌入率（从 1×24 每像素比特数到 6×24 每像素比特数）下的图像质量对比结果。隐写图像的高质量至关重要，因为它能更好地避免人工或机器审查。从图 5(a) 和 5(b) 可以看出，对于基线方法，随着嵌入图像数量的增加，隐写图像的质量逐渐下降，因为为了嵌入更多图像，需要在载体图像中引入更多扰动。相比之下，我们的 LDH 在保持质量方面仍然表现更好

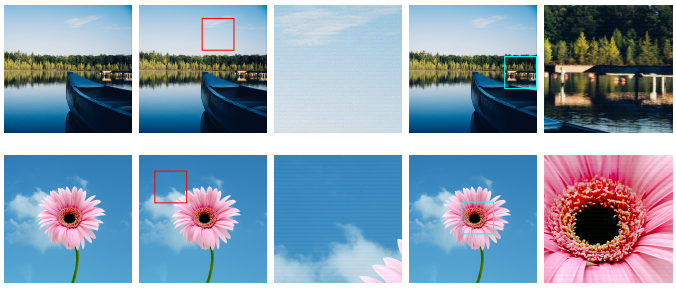


图 4. JND引导的位置与随机位置嵌入秘密代码 S_e 的视觉比较。**第 1 列：**原始载体图像 C。**第 2 列：**通过随机位置嵌入生成的隐写图像 C'：S_e 嵌入红色方框中。**第 3 列：**放大红色方框中的部分。**第 4 列：**在 JND 引导下生成的 C'： S_e 嵌入蓝色方框中。**第 5 列：**放大蓝色方框中的部分。 C' 下方的数字表示与 C 相比的结构相似性指数。

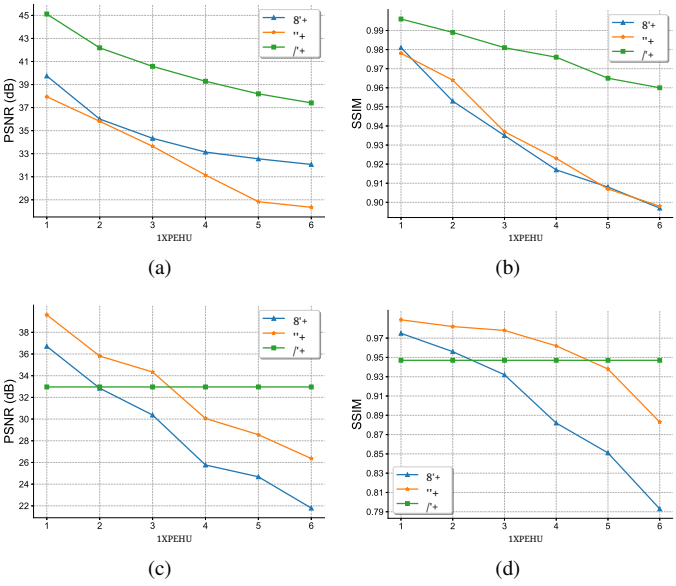


图 5. UDH、DDh 和 LDH 在嵌入不同数量图像时的视觉结果。(a) 载体图像/隐写图像对的峰值信噪比；(b) 载体图像/隐写图像对的结构相似性指数；(c) 秘密图像/恢复图像对的峰值信噪比；(d) 秘密图像/恢复图像对的结构相似性指数。

，尤其是在隐藏多幅秘密图像时，能够更好地维持隐写图像的质量。除了隐写图像的质量外，图 5(c) 和 5(d) 还表明，在 LDH 中，揭示出的秘密图像质量不会受到嵌入图像数量的影响。然而，基线方法对应的视觉质量则会严重下降。综合图 5 可以得出结论：与这些全局方法相比，LDH 更适用于高容量数据隐藏。

C. 超参数的影响

缩放因子。 在我们的 LDH 中，我们使用缩放因子 ω 来确定紧凑秘密代码的大小，这也决定了嵌入率的上限。ω 越大，嵌入率越高。然而，较大的 ω也意味着秘密代码极其紧凑，这不可避免地会导致秘密信息丢失，并使高质量秘密图像难以恢复。如果

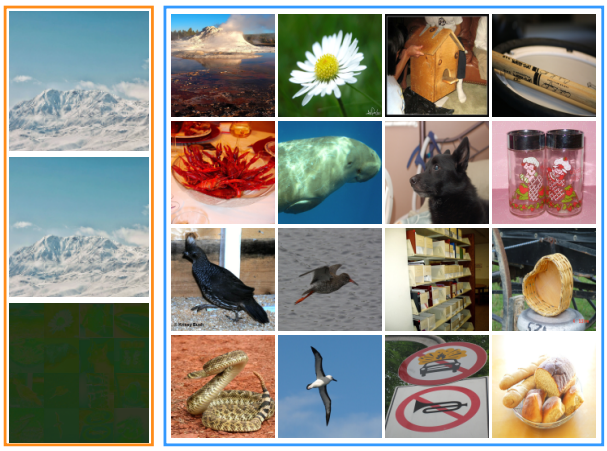


图 6. LDH 的最大嵌入率（即 16×24 每像素比特数）示意图，其中 ω 为 4。从上到下，黄色框中的图像为 C、 C'、 ×15 在 C' 与 C 之间的放大误差。蓝色框中为 16 张提取图像 S'。

恢复出的秘密图像质量过低，相应的隐藏就毫无意义且无法成功。我们在表 II 中证实了这一现象。结合表 II 和表 I 中 LDH 的结果，我们发现 $\omega = 4$ 能够更好地平衡嵌入率与 S' 的质量。随着 ω 从 2 增至 4，隐写图像的质量有所提升，嵌入率达到16×24 每像素比特数，同时 S' 的质量牺牲了约 6.24%。当 $\omega = 4$ 时， S' 的质量可以接受，且原始图像与恢复出的秘密图像之间的峰值信噪比大于 32 分贝。图 6 展示了在 $\omega = 4$条件下 LDH 的视觉效果；LDH 的恢复结果几乎与仅嵌入一张图像时的恢复结果相同，即所有恢复图像的质量都很高。当我们进一步将 ω 从 4 增至 8 时，尽管理论嵌入率可达到 64×24 每像素比特数，但 S' 的质量会变得无法接受（相应的峰值信噪比仅约为 26 分贝）。

模型容量。 在深度隐藏中，深度神经网络具有强大的表征能力，能够将图像信息编码到其参数中，这是深度隐藏相较于传统数据隐藏的最大优势，也是实现高嵌入率的关键。此外，我们认为，容量更大的深度神经网络能够支持更高的嵌入率。为验证这一点，我们测试了多个不同规模的网络。具体而言，我们改变提取网络所有层的宽度，以控制隐藏特征的数量（nhf），并在表 III 中记录相应的实验结果。从表 III 可以看出，较大的 nhf（即更宽的提取网络）会带来更高的峰值信噪比和结构相似性指数。因此，模型容量对于深度隐藏实现高嵌入率至关重要。

D. 鲁棒性分析

在 C' 的传输过程中， C' 可能会受到不可避免的信道噪声或人为扰动（例如 JPEG 压缩、旋转和裁剪）的影响。在这种情况下，强大的鲁棒性能够使隐藏方案在现实世界中更具实用性。接下来，我们将展示，我们的 LDH 能够自然抵御

表 II 不同缩放因子 ω下 LDH 的性能。

数据集	ω=2，1 幅图像						ω=8，1 幅图像					
	载体图像/隐写图像对			秘密/恢复图像对			载体图像/隐写图像对			秘密/恢复图像对		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
COCO	1.321	0.976	38.246	3.527	0.965	33.93	0.114	0.998	48.19	8.606	0.822	25.891
ImageNet	1.36	0.972	38.078	2.676	0.977	36.33	0.117	0.998	48.045	8.078	0.843	26.665
数据集	ω=2，2 张图像						ω=8，2 张图像					
	载体图像/隐写图像对			秘密/恢复图像对			载体图像/隐写图像对			秘密/恢复图像对		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
COCO	2.608	0.947	35.288	3.527	0.965	33.93	0.228	0.996	45.182	8.606	0.822	25.891
ImageNet	2.666	0.949	35.186	2.676	0.977	36.33	0.234	0.996	45.033	8.078	0.843	26.665
数据集	ω=2，4 张图像						ω=8，4 张图像					
	载体图像/隐写图像对			秘密/恢复图像对			载体图像/隐写图像对			秘密/恢复图像对		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
COCO	5.34	0.88	32.141	3.527	0.965	33.93	0.454	0.992	42.177	8.606	0.822	25.891
ImageNet	5.381	0.889	32.118	2.676	0.977	36.33	0.468	0.991	42.028	8.078	0.843	26.665

表 III 不同提取网络容量下揭示图像的图像质量结果。结果通过秘密图像与揭示图像之间的峰值信噪比和结构相似性指数进行衡量。

模型	nhf=48	nhf=64	nhf=80
PSNR	31.382	32.537	33.169
SSIM	0.930	0.942	0.951

表IV DDH、UDH 和我们的 LDH 遭受裁剪攻击后的恢复质量结果。结果以峰值信噪比（PSNR）衡量。后缀数字表示随机裁剪的区块数量。

方法	DDH	UDH	LDH
Crop-1	23.699	21.573	30.863
Cropout-1	23.571	21.557	30.656
Crop-2	20.228	18.918	29.022
Cropout-2	20.217	18.814	29.843
Crop-4	17.225	16.509	26.059
Cropout-4	17.227	16.461	26.255

局部失真，并且可以通过对抗训练进一步增强，以抵御全局失真。

局部失真。 由于采用了特殊的局部嵌入策略，LDH 能够抵御局部失真，例如随机裁剪。现有的全局深度隐藏方案受低冗余这一弱点所限 [64] ，隐写图像中任何部分的缺失都会导致相应秘密像素恢复错误。然而，在我们的 LDH 中，隐写图像随机缺失一部分可能不会影响秘密图像的揭示，因为秘密代码是以局部且随机的方式嵌入的。因此，即使隐写图像遭受了一些局部失真，LDH 仍有可能揭示完整的秘密图像。并且，秘密代码越紧凑（即 ω 越大），完整恢复的概率就越高。

为了展示 LDH 固有的鲁棒性，我们采用裁剪操作，擦除隐写图像中的部分像素。具体而言，裁剪后，我们可以用载体图像的相应部分填充隐写图像中的空洞（*Cropout*），也可以直接用 0 填充空洞（*Crop*）。在图7中，我们使用 Crop 和 Cropout 处理隐写图像（裁剪空洞的大小为 400×400）。从图7中，我们可以直观地看到 LDH 对 Crop 和 Cropout 的固有鲁棒性。如果裁剪区域与嵌入区域不重叠，Crop 和 Cropout 操作不会影响 LDH 中秘密代码的揭示。然而，全局隐藏方案（例如，DDH [16]和 UDH [17]）在裁剪操作后显然无法恢复完整的 S'。为了更好地说明这一情况，我们还在图7中展示了发生重叠的示例（第3列和第7列）。裁剪攻击后的具体峰值信噪比数值列于表IV中。首先，我们将隐写图像划分为 16 个互不重叠的块，每个块包含 256×256 个像素，然后随机选取 1、2 和 4 个块执行 Crop 或 Cropout。从表IV可以清楚地看出，由于揭示的秘密图像质量高得多，LDH 比另外两种全局方案更能抵御裁剪攻击。

全局失真。 除了局部失真之外，我们的 LDH 还可以通过针对性对抗训练 [17]抵御其他常见的全局失真。在对抗训练中，我们在隐藏模块和揭示模块之间添加一个噪声层，用于模拟可能的失真并处理图像 C'。随后，我们训练 P 和 R 识别嵌入区域，并分别从处理后的图像 C' 中揭示所需的秘密图像。具体而言，我们在此考虑三种不同的失真，以验证 LDH 所获得的增强效果：随机失活 （随机失活以一定概率将 C' 中的每个像素替换为 C 中对应的像素。我们将该概率设为 0.3）、高斯模糊 （简称高斯）和*JPEG*压缩 （简称 JPEG）。然后，我们使用随机失活层、高斯层和 JPEG 层作为对抗训练中的噪声层。请注意，我们只能使用三种噪声层中的一种来处理 C' （专用模式），也可以使用全部三种噪声层来处理 C' （组合模式）。

我们评估了基础模型（即未进行对抗训练）和相应增强模型的鲁棒性，并将结果展示在表V中。从表V可以清楚地观察到，在进行对抗训练之前，基础模型面对所有考虑的失真均无法正常工作，因为揭示出的秘密图像质量极低（峰值信噪比$<$ 13 dB），这意味着相应的秘密图像

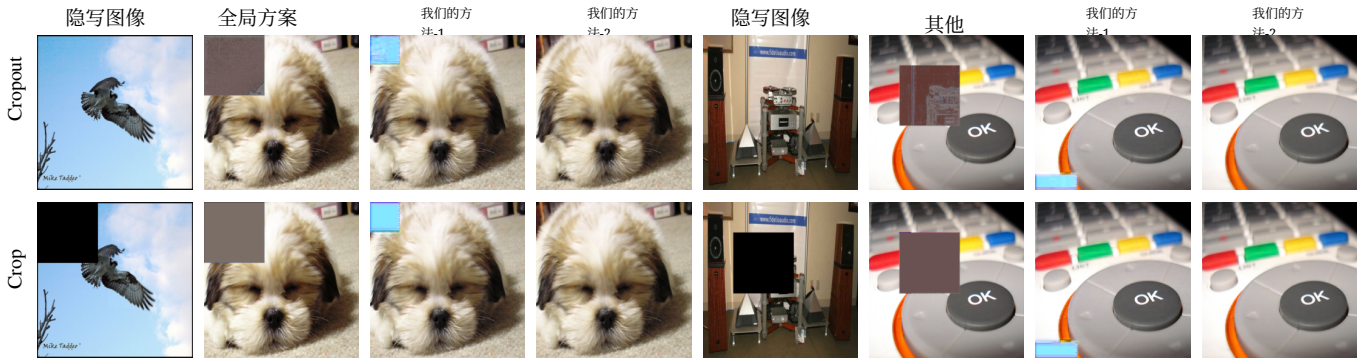


图7. 对隐写图像进行 Cropout 和裁剪攻击后的恢复效果可视化。前四列是在左上角受到攻击，后四列是在隐写图像的随机位置受到攻击。图像尺寸为 1024×1024 像素，裁剪区域的尺寸为 400×400 像素。

表V 不同图像失真下隐写图像和秘密图像的视觉质量结果。“常规训练”：不使用失真的训练；“专门训练”：使用单一对应失真的训练；“组合训练”：使用组合失真的训练。“原始”表示实验过程中未添加图像失真。

失真类型	隐写								
	基础模型			专用模型			组合模型		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
原始	0.343	0.994	43.928	0.418	0.991	42.766	0.583	0.985	40.619
随机失活	0.336	0.993	44.039	0.413	0.991	42.876	0.588	0.987	40.562
高斯	0.344	0.993	43.926	0.474	0.990	41.878	0.586	0.986	40.586
JPEG	0.344	0.994	43.938	0.557	0.987	40.909	0.585	0.986	40.597
失真类型	秘密								
	基础模型			**专用模型**			**组合模型**		
	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑	APD↓	结构相似性指数↑	峰值信噪比↑
原始	3.667	0.957	33.615	4.477	0.934	31.961	9.960	0.834	25.431
随机失活	48.370	0.248	12.420	7.564	0.885	26.947	11.685	0.832	23.230
高斯	52.714	0.309	11.854	5.047	0.917	30.533	9.787	0.856	24.927
JPEG	98.625	0.187	6.659	17.974	0.737	20.447	16.956	0.740	20.711

毫无用处。然而，经过对抗训练后，LDH 的鲁棒性得到了显著提升。对于专用模型，在三种失真下揭示出的秘密图像的峰值信噪比均比基础模型高出2–3倍。此外，组合模型获得的秘密图像质量略低于专用模型，但仍处于可接受水平。组合模型的结果相比基础模型有了显著提升，能够抵抗多种失真，并具有更强的泛化能力。

V. 嵌入率分析

数据隐藏可以视为一个双人博弈：发送方将秘密消息嵌入载体图像中，接收方则从隐写图像中恢复秘密消息。正如我们在表 II 中所讨论的，揭示图像的质量对于数据隐藏至关重要，因为严重失真会使隐藏失去意义。本节将从接收方的角度讨论嵌入率。

在评估实验中，我们发现，恢复的秘密图像（见图 8）中的失真具有特定模式，例如不期望出现的条纹纹理、C 的轮廓纹理以及颜色失真。这些失真可以通过额外的图像复原算法减轻，甚至消除。因此，我们考虑具有不同能力的接收方，以讨论

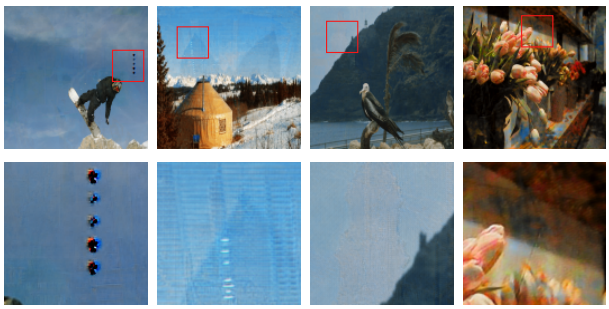


图 8. 隐藏多幅图像时，DDH [16], UDH [17], HCVS [51] 和 MISDNN [53] 的恢复示例（从左到右）。第一行展示揭示的秘密图像，第二行展示方框区域的放大版本。

不同深度隐藏方案的嵌入率上限。具体而言，我们假设接收方可以使用非学习型算法或基于学习的算法，进一步处理揭示出的秘密图像。此外，还需要为恢复的秘密图像质量设定阈值，以讨论嵌入率。我们将深度隐藏方案的最终嵌入率定义为：特定方案在满足该阈值时能够达到的最高嵌入率。

采用非学习型算法的接收方。 我们假设接收方可以采用一些非学习型图像增强或去噪算法，进一步处理这些

表 VI 接收方拥有非学习型图像恢复算法时不同方案的嵌入率。阈值为 32 分贝。

方法	BPDHE [65]	AMSR [66]	LIME [67]	NLFMT [68]	None
DDH [16]	3×24	3×24	3×24	3×24	3×24
UDH [17]	2×24	2×24	2×24	2×24	2×24
MISDNN [53]	1×24	1×24	1×24	1×24	1×24
HCVS [51]	3×24	3×24	3×24	3×24	3×24
LDH	16×24	16×24	16×24	16×24	16×24

表 VII 接收方拥有学习型图像恢复算法时不同方案的嵌入率。

方法	峰值信噪比≥26dB			峰值信噪比≥32dB		
	Restomer	MPRNet	None	Restomer	MPRNet	None
DDH [16]	6×24	5×24	5×24	4×24	3×24	3×24
UDH [17]	5×24	4×24	3×24	3×24	2×24	2×24
MISDNN [53]	4×24	3×24	2×24	1×24	1×24	1×24
HCVS [51]	6×24	5×24	5×24	4×24	3×24	3×24
LDH	64×24	64×24	64×24	16×24	16×24	16×24

恢复图像。这些算法通常较为轻量且易于实现，不要求接收方拥有昂贵的计算资源和数据资源。当我们将阈值设为 32 分贝时，不同方案的实验结果记录在表 VI 中。

采用基于学习的算法的接收方。接收方可能拥有昂贵的计算资源和数据资源，因此能够训练强大的图像复原模型，以提升揭示出的秘密图像的图像质量。在这项实验中，我们考虑了两种最先进的图像复原模型：Restormer [69] 和 MPRNet [70]„ 并针对不同的隐藏方案分别对它们进行训练。我们在表 VII 中展示了高能力接收方的嵌入率。

从表 VI 和表 VII 中，我们观察到，采用非学习型算法的接收方无法提高所涉及方案的嵌入率，而采用基于学习的算法的接收方能够提升基线方法的秘密图像质量。其中，Restomer 对较低和较高阈值均有效，MPRNet 则对较低阈值有效。我们的 LDH 在嵌入率方面具有明显优势：嵌入率很高，且不受接收方采用不同图像恢复方法的影响。实验结果表明，当阈值大于 26 dB 时，所涉及的基线方法最多只能隐藏 6 张图像，而我们的 LDH 可以隐藏 64 张图像。

VI. 结论

本文提出了一种局部深度隐藏（LDH）方案，用于说明深度隐藏在提高嵌入率方面取得成功的原因。LDH利用我们的隐藏网络将秘密图像转换为一段小型、不可感知的紧凑编码，并将该编码嵌入载体图像的随机区域中。利用定位与揭示网络，可以高质量地揭示秘密图像。大量实验表明，所提出的方法不仅具有较高的嵌入率，而且比现有深度隐藏方案实现了更强的鲁棒性。我们还分析了不同图像复原算法对不同深度隐藏方案嵌入率的影响。

参考文献

[1] T. Morkel、J. H. Eloff 和 M. S. Olivier，《图像隐写概述》，载于南非年度信息安全会议，2005年，第1–11页。[2] A. Cheddad、J. Condell、K. Curran 和 P. Mc Kevitt，《数字图像隐写：当前方法的调查与分析》，《信号处理》，第90卷第3期，第727–752页，2010年。[3] R. J. Anderson 和 F. A. Petitcolas，《论隐写术的局限》，《IEEE通信选定领域期刊》，第16卷第4期，第474–481页，1998年。[4] S. K. Bandyopadhyay、D. Bhattacharyya、D. Ganguly、S. Mukherjee 和 P. Das，《隐写术教程式综述》，载于当代计算国际会议，2008 年，第105–114页。[5] M. Liu、W. Luo、P. Zheng 和 J. Huang，《一种用于增强图像隐写的对抗嵌入新方法》，《IEEE信息取证与安全汇刊》，第16卷，第4621–4634页，2021年。[6] Q. Guan、P. Liu、W. Zhang、W. Lu 和 X. Zhang，《利用信道知识的双层双综合征网格码实现鲁棒隐写》，《IEEE信息取证与安全汇刊》，2022年。[7] T. Pevn`y、T. Filler 和 P. Bas，《利用高维图像模型执行高度不可检测的隐写》，载于信息隐藏国际研讨会，2010年，第161–177页。[8] C.-K. Chan 和 L.-M. Cheng，《通过简单的LSB替换在图像中隐藏数据》，《模式识别》，第37卷第3期，第469–474页，2004年。[9] A. A. Tamimi、A. M. Abdalla 和 O. Al-Allaf，《使用可变速率隐写在另一幅图像中隐藏图像》，《国际高级计算机科学与应用期刊》，第4卷第10期，2013年。[10] X. Li 和 J. Wang，《一种基于JPEG和粒子群优化算法的隐写方法》，《信息科学》，第177卷第15期，第3099–3109页，2007年。[11] K. Raja、C. Chowdary、K. Venugopal 和 L. Patnaik，《一种在原始图像上使用LSB、DCT和压缩技术的安全图像隐写方法》，载于智能感知与信息处理国际会议，2005年，第170–176页。[12] R. T. McKeon，《电影中的奇异傅里叶隐写术》，载于IEEE电气/信息技术国际会议，2007年，第178–182页。[13] T. Denemark 和 J. Fridrich，《通过同步选择信道提高隐写安全性》，载于ACM信息隐藏与多媒体安全研讨会，2015年，第5–14页。[14] B. Li、M. Wang、J. Huang 和 X. Li，《一种用于空间图像隐写的新型代价函数》，载于IEEE图像处理国际会议，2014年，第4206–4210页。[15] B. Li、M. Wang、X. Li、S. Tan 和 J. Huang，《空间图像隐写中的修改方向聚类策略》，《IEEE信息取证与安全汇刊》，第10卷第9期，第1905–1917页，2015年。[16] S. Baluja，《藏于众目睽睽之下：深度隐写》，《神经信息处理系统进展》，第30卷，2079，2017年2069页–。[17] C. Zhang、P. Benz、A. Karjauv、G. Sun 和 I. S. Kweon，《UDH：用于隐写、数字水印和光场消息传递的通用深度隐藏》，《神经信息处理系统进展》，第33卷，第10 223–10 234页，2020年。[18] C. Yu，《基于注意力机制的生成对抗网络数据隐藏》，载于AAAI人工智能会议，2020年，第1120–1128页。[19] R. Zhang、S. Dong 和 J. Liu，《通过生成对抗网络实现不可见隐写》，Multimedia Tools and Applications，第78卷第7期，第8559–8575页，2019年。[20] S.-P. Lu、R. Wang、T. Zhong 和 P. L. Rosin，《基于可逆神经网络的大容量图像隐写》，载于IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021年，第10 816–10 825页。[21] M. Tancik、B. Mildenhall 和 R. Ng，《StegaStamp：实体照片中的不可见超链接》，载于IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020年，第2117–2126页。[22] E. Wengrowski 和 K. Dana，《利用深度摄影隐写实现光场消息传递》，载于IEEE/CVF Conference on Computer Vision and Pattern Recognition，2019年，第1515–1524页。[23] N. Provos 和 P. Honeyman，《藏与寻：隐写术入门》，《IEEE安全与隐私》，第1卷第3期，第32–44页，2003年。[24] A. M. Fard、M.-R. Akbarzadeh-T 和 F. Varasteh-A，《一种用于安全JPEG隐写的新型遗传算法方法》，载于IEEE智能系统工程国际会议

，2006，第 1–6 页。

[25] A. A. Attaby、M. F. M. Ahmed 和 A. K. Alsammak，“使用一种新颖技术在 JPEG 图像中隐藏数据并增强其抵抗隐写分析的能力：DCT-M3”，Ain Shams Engineering Journal，第 9 卷，第 4 期，第 1965–1974 页，2018年。[26] H. M. Reddy 和 K. Raja，“使用离散小波变换实现高容量和高安全性的隐写术”，International Journal of Computer Science and Security，第 3 卷，第 6 期，第 462 页，2009年。[27] S. Atawneh、A. Almomani、H. Al Bazar、P. Sumari 和 B. Gupta，“基于 DWT 域菱形编码的安全且不可感知数字图像隐写算法”，Multimedia Tools and Applications，第 76 卷，第 18 期，第 18 451–18 472 页，2017年。[28] J. Ruanaidh、W. Dowling 和 F. M. Boland，“数字图像的相位水印”，载于 IEEE图像处理国际会议，第 3 卷，1996年，第 239–242 页。[29] A. K. Gulve 和 M. S. Joshi，“一种使用像素值差分方法将秘密数据隐藏于整数小波变换系数中的图像隐写方法”，Mathematical Problems in Engineering，第 2015 卷，2015年。[30] M. Kalita、T. Tuithung 和 S. Majumder，“一种使用整数小波变换和最低有效位替换的新型隐写方法”，The Computer Journal，第 62 卷，第 11 期，第 1639–1655 页，2019 年。[31] K.-H. Jung 和 K.-Y. Yoo，“ 基于数字图像插值与最低有效位替换的隐写方法”，Multimedia Tools and Applications，第 74 卷，第 6 期，第 2143–2155 页，2015年。[32] A. K. Sahu 和 G. Swain，“使用双层最低有效位匹配的可逆图像隐写”，Sensing and Imaging，第 21 卷，21，2020 1，1第。页 – 。[33] P. C. Mandal 和 I. Mukherjee，“基于多方向像素值差异和减小差值扩展的高容量数据隐藏”，Multimedia Tools and Applications，第 81 卷，第 4 期，第 5325–5347 页，2022年。[34] Z. Li 和 Y. He，“基于粒子群优化的像素值差分与模函数隐写术”，Journal of information security and applications，第 43 卷，第 47–52 页，2018年。[35] D. M. Thodi 和 J. J. Rodríguez，“可逆水印的扩展嵌入技术”，IEEE Transactions on Image Processing，第 16 卷，第 3 期，第 721–730 页，2007年。[36] W.-L. Tai、C.-M. Yeh 和 C.-C. Chang，“基于像素差直方图修改的可逆数据隐藏”，IEEE视频技术电路与系统汇刊，第 19 卷，910，2009 6，906第。页 – 。[37] X.-Z. Xie、C.-C. Chang 和 Y.-C. Hu，“一种利用有符号数字表示、基于预测误差直方图移位的自适应可逆数据隐藏方案”，Multimedia Tools and Applications，第 79 卷，第 33 期，第 24 329–24 346 页，2020年。[38] J. Tian，“使用差值扩展进行可逆数据嵌入”，IEEE Transactions on Circuits and Systems for Video Technology，第 13 卷，第 8 期，第 890–896 页，2003年。[39] S. Gujjunoori 和 M. Oruganti，“基于差值扩展的可逆数据嵌入与边缘检测”，Multimedia Tools and Applications，第 78 卷，第 18 期，第 25 889–25 917 页，2019年。[40] T. D. Kieu 和 C.-C. Chang，“一种采用多重嵌入策略、兼具高隐写图像质量、可逆性和高载荷的隐写方案”，Journal of Systems and Software，第 82 卷，第 10 期，第 1743–1752 页，2009年。[41] H. Nyeem，“采用图像位平面切片的可逆数据隐藏”，载于计算机与信息技术国际会议，2017年，第 1–6 页。[42] J. Fridrich、M. Goljan 和 R. Du，“可靠检测彩色和灰度图像中的最低有效位隐写”，载于多媒体与安全研讨会：新挑战，2001年，第 27–30 页。[43] S. Dumitrescu、X. Wu 和 Z. Wang，“通过样本对分析检测最低有效位隐写”，载于信息隐藏国际研讨会，2002年，第 355–372 页。[44] J. Fridrich 和 M. Goljan，“对空域最低有效位隐写中秘密消息长度的估计”，载于多媒体内容安全、隐写与水印 VI，第 5306 卷，2004年，第 23–34 页。[45] A. D. Ker，“改进灰度图像中最低有效位隐写的检测”，载于信息隐藏国际研讨会，2004年，第 97–115 页。[46] ——，“最低有效位替换结构隐写分析的通用框架”，载于信息隐藏国际研讨会，2005年，第 296–311 页。[47] J. Hayes 和 G. Danezis，“通过对抗训练生成隐写图像”，《神经信息处理系统进展》，第 30 卷，2017年。[48] S. Ghamizi、M. Cordy、M. Papadakis 和 Y. L. Traon，“对抗嵌入：一种稳健且难以察觉的隐写与水印技术”，arXiv 预印本 arXiv:1912.01487，2019年。[49] D. Hu、L. Wang、W. Jiang、S. Zheng 和 B. Li，“一种基于深度卷积生成对抗网络的新型图像隐写方法”，IEEE Access，第 6 卷，第 38 303–38 314 页，2018年。[50] B. Ray、S. Mukhopadhyay、S. Hossain、S. K. Ghosal 和 R. Sarkar，“基于深度学习的边缘检测图像隐写”，《多媒体工具与应用》，第 80 卷，第 24 期，第 33 475–33 503 页，2021年。[51] X. Weng、Y. Li、L. Chi 和 Y. Mu，“基于时间残差建模的高容量卷积视频隐写”，载于 International Conference on Multimedia Retrieval，2019年，第 87–95 页。[52] R. Rahim、S. Nadeem 等，“端到端训练的卷积神经网络编码器-解码器网络用于图像隐写”，载于 European Conference on Computer Vision Workshops，2018年，第 0–0 页。[53] A. Das、J. S. Wahi、M. Anand 和 Y. Rana，“使用深度神经网络的多图像隐写”，arXiv 预印本 arXiv:2101.00350，2021年。[54] K. A. Zhang、A. Cuesta-Infante、L. Xu 和 K. Veeramachaneni，“Steganogan：基于 GAN 的高容量图像隐写”，arXiv 预印本 arXiv:1901.03892，2019年。[55] R. Meng、Q. Cui、Z. Zhou、Z. Fu 和 X. Sun，“一种基于 CycleGAN、用于物联网隐蔽通信的隐写算法”，IEEE Access，第 7 卷，第 90 574–90 584 页，2019年。[56] Z. Guan、J. Jing、X. Deng、M. Xu、L. Jiang、Z. Zhang 和 Y. Li，“DeepMIH：用于多图像隐藏的深度可逆网络”，IEEE Transactions on Pattern Analysis and Machine Intelligence，2022年。[57] P. Isola、J.-Y. Zhu、T. Zhou 和 A. A. Efros，“使用条件对抗网络进行图像到图像转换”，载于 IEEE Conference on Computer Vision and Pattern Recognition，2017年，第 1125–1134 页。[58] L. Dinh、D. Krueger 和 Y. Bengio，“NICE：非线性独立成分估计”，arXiv 预印本 arXiv:1410.8516，2014年。[59] L. Dinh、J. Sohl-Dickstein 和 S. Bengio，“使用 Real NVP 进行密度估计”，arXiv 预印本 arXiv:1605.08803，2016年。[60] J. Wu、L. Li、W. Dong、G. Shi、W. Lin 和 C.-C. J. Kuo，“针对具有图案复杂度的图像改进恰可察觉差异模型”，IEEE Transactions on Image Processing，第 26 卷，第 2693 页，2017年 6，第 2682 页，第 期。[61] O. Ronneberger、P. Fischer 和 T. Brox，“U-Net：用于生物医学图像分割的卷积网络”，载于 International Conference on Medical Image Computing and Computer-assisted Intervention，2015年，第 234–241 页。[62] C. Dong、C. C. Loy 和 X. Tang，“加速超分辨率卷积神经网络”，载于 European Conference on Computer Vision，2016年，第 391–407 页。[63] D. P. Kingma 和 J. Ba，“Adam：一种随机优化方法”，arXiv 预印本 arXiv:1412.6980，2014年。[64] T. Xiang、H. Liu、S. Guo 和 T. Zhang，“PEEL：一种针对深度隐藏的可证明移除攻击”，arXiv 预印本 arXiv:2106.02779，2021年。[65] H. Ibrahim 和 N. S. P. Kong，“用于图像对比度增强的保亮度动态直方图均衡化”，IEEE Transactions on Consumer Electronics，第 53 卷，第 4 期，第 1752–1758 页，2007年。[66] C.-H. Lee、J.-L. Shih、C.-C. Lien 和 C.-C. Han，“用于图像对比度增强的自适应多尺度 Retinex”，载于 International Conference on Signal-Image Technology & Internet-Based Systems，2013年，第 43–50 页。[67] X. Guo，“LIME：一种低照度图像增强方法”，载于 ACM International Conference on Multimedia，2016年，第 87–91 页。[68] B. Shreyamsha Kumar，“基于非局部均值滤波器及其方法噪声阈值处理的图像去噪”，Signal, Image and Video Processing，第 7 卷，第 6 期，第 1211–1227 页，2013年。[69] S. W. Zamir、A. Arora、S. Khan、M. Hayat、F. S. Khan 和 M.-H. Yang，“Restormer：用于高分辨率图像复原的高效 Transformer”，载于 IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022年，第 5728–5739 页。[70] S. W. Zamir、A. Arora、S. Khan、M. Hayat、F. S. Khan、M.-H. Yang 和 L. Shao，“多阶段渐进式图像复原”，载于 IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021年，第 14 821–14 831 页。