

# 面向无载体信息隐藏的映射关系智能搜索方法

王亚宁<sup>1,2</sup>, 吴 彬<sup>1,2</sup>

<sup>1</sup> 中国科学院信息工程研究所信息安全国家重点实验室 北京 中国 100093

<sup>2</sup> 中国科学院大学 网络空间安全学院 北京 中国 100049

**摘要** 传统的搜索式无载体信息隐藏技术建立在固定的映射规则与庞大的图像库基础上, 依赖于复杂的人工特征提取并且需要进行大量搜索来构建合适的图像库。针对这些问题, 本文提出了一种面向无载体信息隐藏的基于深度学习映射关系智能搜索方法, 该方法从已有图像库出发, 基于深度神经网络, 自动搜索一套高容量、高覆盖率的映射关系, 从而解决传统人工方法存在的传输开销大、图像库建立困难的问题。除此以外, 实验表明我们的方法相较于传统无载体方法有更强的鲁棒性。

**关键词** 信息隐藏; 隐写术; 无载体信息隐藏; 深度学习

**中图分类号** TN915.08 **DOI号** 10.19363/J.cnki.cn10-1380/tn.2020.05.05

## An intelligent search method of mapping relation for coverless information hiding

WANG Yaning<sup>1,2</sup>, WU Bin<sup>1,2</sup>

<sup>1</sup> State Key Laboratory of Information, Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

<sup>2</sup> School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** Based on fixed mapping rules and large image database, traditional searching coverless information hiding technology relies on complex artificial feature extraction and takes a lot of searching to construct available image database. To solve these problems, this paper proposed an intelligent search method of mapping relation for coverless information hiding based on deep learning. This framework based on the collected image set and automatically searched for a set of high-capacity, high-coverage mapping rules based on deep neural networks, thereby solving the problems: large transmission overhead and a large image database of traditional coverless methods. In addition, experiments show that our method is more robust than the traditional coverless methods.

**Key words** information hiding; steganography; coverless steganography; deep learning

### 1 引言

在互联网时代, 对网络中传输的各种信息进行攻击和截获变得更为容易, 一旦用户的重要信息被监听或破坏, 将会带来不可估量的损失。传统的密码学算法虽然可以将重要信息加密, 以实现网络中的保密传输, 但是加密后的信息往往以乱码形式存在, 在某些情况下, 该种形式更容易引起攻击者的怀疑和关注。

为了实现更高级别的安全通信, 在特定的环境下可以采取隐蔽通信技术。Shannon<sup>[1]</sup>将信息安全系统归结为三种: 隐蔽通信系统, 加密系统和隐私系统。其中隐蔽通信系统指的是通信方可以将秘密信

息以第三方不可感知的方式隐藏在载体中, 并将载体以公开的方式进行发布或传输, 而不会引起监控方的注意和怀疑。

如图 1 所示, 在 Simmons<sup>[2]</sup>提出的经典隐蔽通信模型中涉及三方, 包括 Alice, Bob 和 Eve。Alice 首先将秘密消息 secret 隐藏在载体 cover 中, 从而得到载密文件 container, 之后将 container 发送给 Bob。但是 Alice 并不是每次都发送载密文件, 有时也会发送未载密的载体 cover 给 Bob。Eve 是一个隐写分析者, 他的目标是判断 Alice 和 Bob 之间传递的文件是否包含了秘密消息, 也就是区分信道中传输是载密文件 container 还是载体 cover。Eve 只需做出判断, 无需解码出对应的秘密消息。在该框架下, Alice 成功进行

**通讯作者:** 吴彬, 博士, 副研究员, Email: wubin@iie.ac.cn。

本课题得到国家自然科学基金(No.U1936119, No.61941116)和国家重点研发计划课题(GrantNo.2019QY(Y)0602)资助。

收稿日期: 2020-01-25; 修改日期: 2020-05-03; 定稿日期: 2020-05-06

秘密消息传输的条件如下: 1) Bob 使用解码协议可以成功从 container 中解码出秘密消息; 2) Eve 判定传输的文件是 container 还是 cover 的几率为 50%。该过程与深度学习中对抗训练的期望目标相似<sup>[3]</sup>。

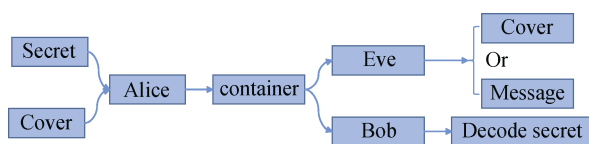


图 1 隐蔽通信三方模型

Figure 1 Hidden communication tripartite model

传统的信息隐藏方法主要利用多媒体数据(载体)的冗余性, 按照一定的规则对其修改, 从而将秘密信息以“不可见”的方式嵌入到载体中<sup>[3]</sup>。然而, 从理论上讲, 只要对载体进行了修改, 那么必然会留下修改痕迹, 就有可能被隐写分析算法检测出来。为了从根本上抵抗隐写分析检测, 近年来信息隐藏领域的研究者提出了“无载体信息隐藏”这一概念。这里的“无载体”指的并不是不需要载体, 而是与传统基于修改的信息隐藏技术相比, 该类方法强调的是不需要修改载体, 而是直接以秘密信息为驱动来“获取/生成”含密多媒体载体<sup>[4]</sup>。

在无载体信息隐藏中, 目前研究较多的是基于关系映射的图像搜索式方法。该类方法首先要定义一套图像到码字的映射规则。然后根据规则去搜索大量的自然图像, 并进行逐一的匹配和筛选, 以获取能够组成一套完备码字的图像库。最终利用这些选中的图像库进行秘密信息传递。该类方法虽然可以从根本上抵抗各类隐写分析工具的检测, 但是它们难以实际应用。限制其发展的原因有两个: 1) 每张图像所能包含的信息容量过小; 2) 需要通过大规模搜索的方式建立庞大的图像库, 而建立的过程中需要耗费大量的时间和计算资源。

为了解决以上图像搜索式无载体信息隐藏存在的问题, 本文抛弃了先设计规则再依照规则搜索图片的思路, 而是提出了一种先给定图像库, 再根据已有图像库自动搜索无载体映射关系的思路。本文所提的方法基于深度学习网络进行映射关系的自动化智能搜索, 可以对用户指定的图像库找到一套高容量、高覆盖率的映射规则。此外通过实验证明本方法还有着优良的鲁棒性以及抗检测性。

本文方法的简单过程如下: 首先为收集到的数据图片指定码字, 接着将码字和图片放入神经网络进行端到端训练, 利用梯度下降和反向传播算法使网络能够学习到每个图像的标签, 然后保存神经网络的模型参数, 得到一个包含映射规则的神经网络

模型。接收方在接收到载密图片之后, 将图片输入相同的神经网络中进行预测输出, 得到秘密消息, 实现消息的提取。与之前方法相比, 本文主要贡献可以归纳为以下三点:

(1) 抛弃了先设计规则, 再根据规则去大规模搜索合适图片的思路, 创新性地提出了先确定图像库, 然后利用深度学习的学习特性对图像库进行映射关系快速智能搜索的框架。

(2) 在所提出的深度学习框架下, 测试了多种经典神经网络结构的表现性能, 实际性能证明本文方法在多个指标上优于传统的基于人工设计规则的图像搜索式无载体方法。

(3) 解决了现有搜索式方法依赖万张级别的大型图像库、隐藏信息的相对容量过小、筛选和搜索过程资源消耗过大的问题, 通过映射关系快速智能搜索方法有效降低了通信初始建立的时间代价, 并且具有覆盖率高、隐藏容量大、抗噪性强、隐蔽性好的特点。

本文内容安排如下: 第 2 节概述了无载体信息隐藏领域的相关工作, 包括近几年提出的图像搜索式方法、基于图像基元的合成方法和基于 GAN 的生成式无载体方法; 第 3 节详细讲述了本文方法的流程和框架, 包括网络的训练与秘密消息隐藏和提取的细节; 第 4 节为实验结果与分析, 从容量、隐蔽性、鲁棒性、覆盖率等方面对算法性能进行了验证; 第 5 节对全文进行总结; 第 6 节分析了现存的问题和未来的工作展望。

## 2 相关工作

本节将介绍最近几年国内外无载体信息隐藏技术的相关工作。

### 2.1 基于 GAN 的完全生成式方法

深度学习是机器学习领域中的新兴方向, 旨在通过模拟人脑自动地学习数据各个层次的抽象特征, 从而更好地反映数据的本质特征。自 2006 年, Hinton<sup>[5]</sup>提出一种基于概率图模型的多层受限波尔兹曼机(Restricted Boltzmann Machine, RBM)后, 深度学习已成为图像处理和计算机视觉领域的主导工具。近年来, 深度学习在图像处理、自然语言处理和语音识别等领域取得了一系列突破性进展, 已经成为学术界的研究热点, 特别是卷积神经网络(Convolutional Neural Network, CNN)<sup>[6]</sup>、深度置信网络(Deep Belief Network, DBN)<sup>[7]</sup>、层叠自动编码器(Stacked Auto-Encoder, SAE)<sup>[8]</sup>、长短期记忆网络(Long-Short Term Memory)<sup>[9]</sup>、生成对抗网络

(Generative Adversarial Network, GAN)等深度模型在各领域推动了突破性成果的大量涌现。

在各种深度网络模型中, GAN 是一种利用数据驱动构造生成模型的方法, 其根本目标是构造一个生成器  $G$ , 向生成器  $G$  输入噪声  $Z$ , 输出生成样本  $G(Z)$ , 且要求  $G(Z)$  服从的分布  $P_g$  与真实数据  $X$  的分布  $P_r$  相同。生成器  $G$  可以看作是一个变换, 它将一个随机噪声  $Z$  转换到真实数据  $X$  的空间中。为了得到这样的一个生成器  $G$ , Goodfellow<sup>[3]</sup>等引入一个判别器  $D$ , 通过不断区分生成样本与真实数据  $X$ , 逐渐改进生成器的性能。

现今已有一些学者将 GAN 应用在无载体信息隐藏领域, 称为基于 GAN 的完全生成式方法。该类方法的基本思想是在图像生成过程中隐藏信息。其通过利用 GAN 强大的生成能力, 将秘密消息作为模型的输入, 生成包含秘密消息的载体。

其中典型的方法有: 刘明明等人<sup>[10]</sup>摒弃了传统的图像合成方法, 首次提出基于 GAN 的生成式无载体图像信息隐藏算法。接着, Duan 等人<sup>[11]</sup>也使用 GAN 来生成图像并隐藏信息; Chu 等<sup>[12]</sup>提出一种基于 CycleGAN 的信息隐藏方法, 实现了在航拍图像中隐藏消息; 之后, Liu 等<sup>[13]</sup>提出基于约束采样的生成式隐写框架。该框架首先训练一个生成器, 然后对生成器进行约束采样, 以获得满足条件的含密载体; Hu 等<sup>[14]</sup>在 DCGAN 的基础上, 提出了一种无载体隐写方法。

## 2.2 基于图像基元合成的半生成式方法

Otori 等人<sup>[15]</sup>最早提出了一种基于像素生成的信息隐藏算法。Xu 等人<sup>[16]</sup>受到湿拓画算法<sup>[17]</sup>的启发, 提出一种基于可逆形变的图像信息隐藏算法。Qian 等人<sup>[18]</sup>则提出一种水影画的生成式信息隐藏算法。另外, Wu 等人<sup>[19]</sup>提出一种基于块的图像信息隐藏, 把具有与秘密信息相同序号的候选块合成到重叠区域来隐藏信息。Zhou 等人<sup>[20]</sup>提出一种基于种子区域生长和最低有效位的纹理合成信息隐藏算法。其使用种子区域生长算法来确定纹理合成的顺序以改掩盖效果不足问题。为解决鲁棒性不足问题, Qin 等人<sup>[21]</sup>摒弃了重叠区域, 而利用图像块的中心区域来隐藏信息。通过中心块的像素均值构造特征。Wei 等人<sup>[22]</sup>不使用中心区域的像素均值构造特征, 而是将均值作为特征输入到 SVM(Support Vector Machine)分类器中进行分类。最后, 使用不同的类别来映射信息。Qian 等人<sup>[23]</sup>也使用中心区域隐藏信息。不同的是其使用中心区域像素的标准差来映射秘密信息。而 Lee

等人<sup>[24]</sup>提出了一种基于样式合成的无载体图像信息隐藏。其选择不同样式作为合成对象, 通过改变它们的颜色、大小和位置三种属性来隐藏信息。近期, Zhang<sup>[25]</sup>等人将分形图像与信息隐藏相结合, 提出可以在分形图形生成过程中通过控制像素渲染来隐藏信息, 该方法有较好的鲁棒性和较大容量, 为基于图像基元合成的半生成式方法提供了新思路。

## 2.3 基于关系映射的图像搜索式方法

基于关系映射的方法通过利用秘密信息和对应图像之间的关系来选择适合图像进行隐藏信息。Fridrich 等<sup>[26]</sup>最早提出了基于载体选择的信息隐藏思路, 其根据一定的规则从事先建立的图像库中检索特定的自然图像来表达秘密信息。目前此类方法主要通过优化图像到码字的映射规则, 来提高方法的性能。如 Zhou 等人<sup>[27]</sup>提出了一种基于像素均值的无载体图像信息隐藏算法: 通过计算像素均值, 为每张图像构建长度为 8 的码字。然而, 该方法没有给出足够的实验分析来验证算法的其他性能。在 Zhou 的基础上, Cao 等人<sup>[28]</sup>引入了 BOW(Bag of Words)算法<sup>[29]</sup>, 大大提升了隐藏的容量, 但是没有过多讨论方法的鲁棒性。为提升方法抗噪声攻击的能力, 有学者使用 SIFT(Local Scale-Invariant Features)<sup>[30]</sup>来构建特征序列<sup>[31, 28]</sup>。Zhou 等人<sup>[32]</sup>利用图像子块检索的方法来隐藏信息。其使用载体图像的子块直接映射秘密信息的子块, 因此秘密信息必须为图像形式。此外 Zou 等人<sup>[33]</sup>提出一种基于子图像像素均值的无载体信息隐藏方案。其将像素均值的取值划分多个区间, 使用区间来代表秘密信息。Zhang 等人<sup>[35]</sup>为了增加无载体图像信息隐藏算法的鲁棒性, 将特征引入到频域, 使用 DC 系数(direct current coefficients)构造特征。Wu 等人<sup>[36]</sup>为进一步提高鲁棒性, 提出了一种基于灰度梯度共生矩阵的无载体图像信息隐藏算法。该算法利用灰度梯度共生矩阵来提取特征值, 能够较好的抵抗旋转、JPEG 压缩攻击和低通滤波攻击。Cao 等人<sup>[37]</sup>则在前人的工作基础上提出了一个无载体动态内容选择框架。最近, Zhou 等人<sup>[38]</sup>利用深度神经网络通过检测并定位图像中的物体对象, 并利用这些对象的标签来表达秘密信息。

然而, 上述方法中的大多数都需要依赖人工方式设计复杂的映射规则。由于每张图片对应的码字不可控, 以上方法都需要搜索大规模的图像库。另一方面, 该类方法的容量较低, 导致发送秘密消息时所需的载体文件开销过大。目前这两个难题阻碍了搜索式无载体的进一步发展。

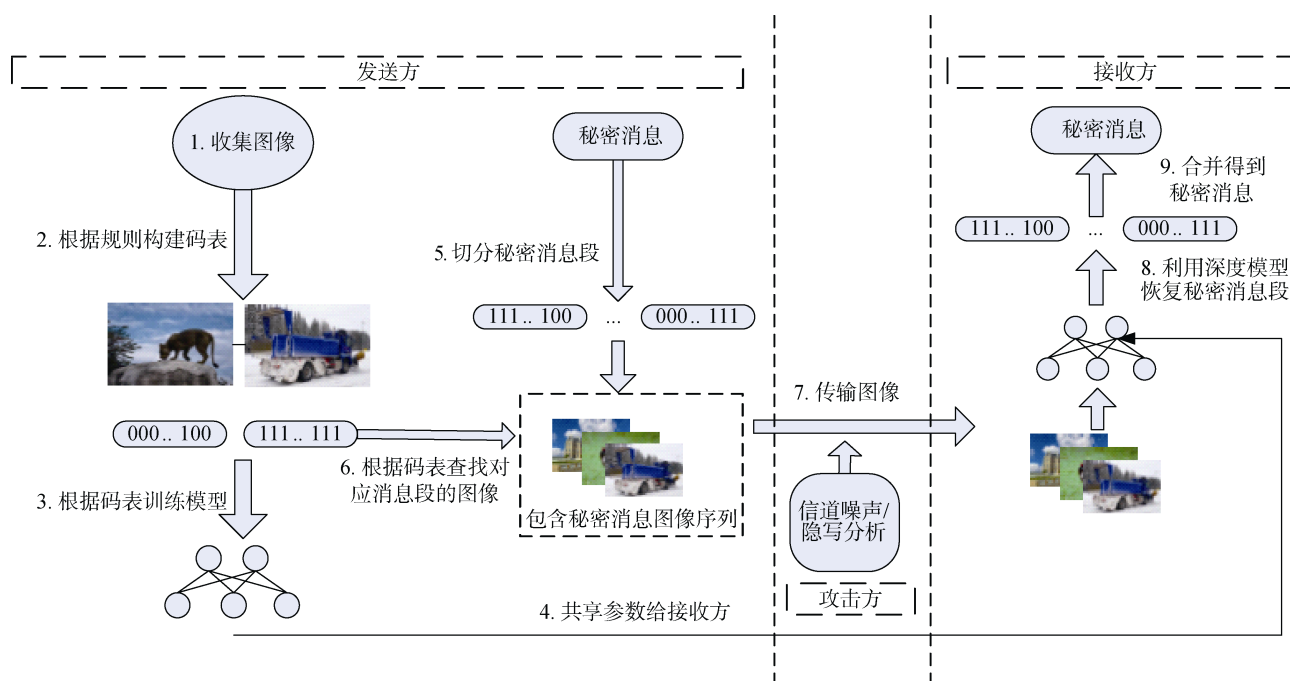


图 2 方法流程图

Figure 2 Flowchart of method

综合来看,在生成式信息隐藏中已经有了很多基于深度学习的工作,而在搜索式的无载体信息隐藏研究中目前尚无利用深度学习技术进行映射规则搜索的工作。我们希望通过引入深度学习技术来解决目前搜索式信息隐藏中存在的问题。

### 3 基于深度学习的无载体映射关系自动搜索方案

本节首先介绍方案的整体流程,之后将从构建码表、数据预处理与模型训练、载密图像序列生成、隐蔽通信与第三方攻击、抽取秘密消息五个关键步骤进行详细介绍。

#### 3.1 方案总体流程

利用本方案进行通信的流程图如图 2 所示。首先,进行的是协商过程:1)发送方收集一定数量的图像构成图像库;2)发送方(Alice)和接收方(Bob)需要根据规则构建一套码表,码表包含了每个码字与载密图片的对应规则。码表可以使发送方和接收方就每张图片代表的秘密消息达成共识,防止出现歧义;3)由发送方根据码表中图像与码字对应关系,训练一个神经网络模型。该模型的训练目标是将给定的输入图像输出为指定的码字;4)达到训练目标后,Alice 将网络的权重参数通过安全方式与 Bob 进行共享,Bob 根据权重参数初始化自己的网络结构,从而得到一个完整的神经网络模型用于解码。至此双方协商过程完成。

双方完成协商之后是隐蔽通信的过程:5)当有发送秘密消息的需求时,发送方首先将秘密消息转换、切分为等长的二进制码段;6)发送方根据索引为每个码段找到对应的图片,形成一串包含秘密消息的图像序列;7)发送方将图像序列发送到接收方,图像传输过程中,会受到攻击方的噪声攻击或隐写分析攻击;8)接收方收到图像后,利用神经网络模型恢复秘密消息段;9)接收方将秘密消息段合并解码为可读消息。至此隐蔽通信完成。

#### 3.2 构建码表

在利用神经网络搜索映射关系之前,应首先向图像分配码字。图像码表是图像与码字之间的映射关系表。图像码表的建立是为了方便神经网络进行有监督的学习。为了充分利用图像库的编码潜力,所以本文所提出的自动分配算法遵循以下原则:必须保证一个图像只能映射到一个码字,一个码字应该映射到尽可能少的图像,但至少应该有一个图像对应。这样我们就可以最高效地挖掘已有图像库的编码潜力。

在分配码字时候,首先发送者应该根据图像的数量定义码字的范围:假设我们有  $M$  张图像,最大的编码范围是  $N$  位编码,那么  $N$  和  $M$  满足如下关系:

$$N = \text{Floor}(\log_2^M) \quad (1)$$

例如,如果我们只有  $M=300$  个图像,根据公式我们最多可以为这些图像分配长度为  $N=8$  的编码



(00000000-11111111)。其次, 发送方应保证从 0 到  $2^N-1$  的每个码字都至少映射到一个图像。本文所用的码字分配算法如表 1 所示, 首先对图像进行随机排序, 然后按顺序向每个图像分配一个对应码字, 最后多余的图像从  $2^N$  个码字中随机选择并与之对应。分配完码字之后我们会得到一个码表。其内容包括码字, 以及该码字对应的图像。该码表方便发送方训练模型以及在发送秘密消息时能快速检索到相应的图像。

表 1 码字分配算法

Table 1 Code word assignment algorithm

Algorithm 1 : Assign Codes to Images	
1:	Input: Images $pic=sorted[pic_1, pic_2, \dots, pic_{end}]$
2:	Output: Image-Code Dictionary
3:	Begin:
4:	Enum = getNum(pic)
5:	$N = \text{Floor}(\text{Log}_2^{enum})$
6:	For $i=1$ to $2^N$ do:
7:	Image_Code{pic; $i$ }
8:	End for
9:	For $I = 2^N$ to Enum:
10:	Id = Randomint( $1, 2^N$ )
11:	Image_Code{pic; $I$ }
12:	End for
13:	Return Image_Code
14:	End

本文构造的码表如图 3 所示, 码字长度为 8, 码表总长度为 256, 每个码字可以描述为 8 位的二进制字符串 {00000000, ..., 11111111}。Lena 图像被分配的码字是“10001000”, 那么该图像被放置到码表“10001000”对应的位置。其余所有图像按照相同的步骤填入码表, 在保证每个码字都至少有一张对应图片, 此时形成一个倒排索引。倒排索引方便发送秘密消息时快速找到秘密消息段对应的图像。例如当发送消息包含码字“10001000”时, 可以选择发送一张 Lena 图片。

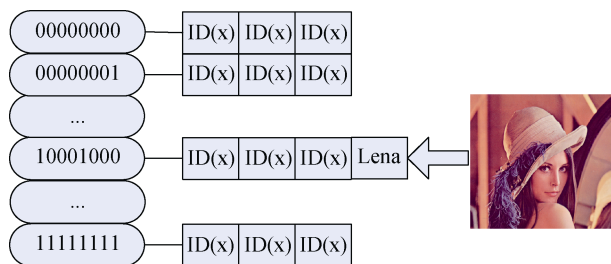


图 3 倒排索引示意图

Figure 3 Inverted index diagram

### 3.3 数据预处理与模型训练

由于深度学习需要大量的数据进行学习, 因此要准备充分的数据对模型进行训练。但是, 根据表 1 中的编码规则, 有时只有一个图片被分配给一个码字, 因此需要一些数据增强方法来提高模型的学习能力。在实验中, 我们采用了如下的图像预处理方法: 首先分别从四个角和中心裁剪给定的图像, 加上这些图像的翻转版本(使用水平翻转)。这样, 一个原始样本图像可以处理成 10 个样本。

由于大图片会大大增加模型训练的时间, 因此本文采用最近邻线性插值的方法将图片大小调整到合适的尺寸, 以加速训练过程。神经网络的损失函数采用了经典的 softmax 函数:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \text{ for } i = 1, \dots, K \quad (2)$$

$$\text{and } z = (z_1, \dots, z_K) \in R^K$$

其中  $z_i$  表示输出属于某一类的概率,  $K=2^{n-1}$ 。

在完成码字分配任务后, 每个图像都有对应的码字。码字类似于图像分类中的标签。然后, 将已经指定好码字的所有图像输入神经网络进行训练。输入为图像, 输出为预测的码字。将预测码字与分配码字进行对比, 并求损失函数以及反向传播, 以使得神经网络的输出尽量靠近给定的标签, 直到能够将所有给定的  $M$  张图片的模型输出全部等于指定码字分为止。其过程类似于训练一个过拟合的图像分类器。最终实现输入一张图像, 输出为码表中对应的标签时, 视为训练完成。

如图 2 所示, 训练完成后, 模型的权重参数被共享给接收方, 接收方用得到的参数初始化模型。

### 3.4 载密图像序列生成

建立码表与训练模型完成后, 现在可以开始载密图像序列构建的过程, 并且该过程不需要修改图像内容。当发送方有发送秘密消息的需求时, 首先将秘密消息转化为二进制码流然后按照一定长度将比特流划分为比特段。得到二进制码段后, 可以通过检索图像码表, 根据 3.2 中构建的倒排索引找到每个比特段对应的图片。例如某比特段为 0100100110001000, 可以选取图 3 所示的索引结构中“01001001”和“10001000”位置对应的图像来代表秘密信息。依此规则得到一串图片序列。此时秘密消息已经隐藏在图片序列中, 发送方只需将图片序列发送给接收方。

### 3.5 隐蔽通信与第三方攻击

载密图像序列构建完成后, 发送方把图像序列发送给接收方。图像序列在信道中传输会遭遇到第三方的攻击和信道噪声的干扰。常见的攻击和噪声干扰包括图像压缩, 缩放, 裁剪, 滤波器攻击, 高斯噪声干扰以及隐写分析等。

一个安全的隐蔽通信模型在面对攻击的时候要求具有: 1)抗检测性: 攻击方只有 50%几率可以正确判断图像是否含有秘密消息; 2)鲁棒性: 接收方收到被干扰后的图像仍然能正确恢复出秘密消息。我们将在第四部分的实验中证明, 我们的方法在抗检测性和鲁棒性上均表现优良。

### 3.6 抽取秘密消息

接收方收到包含秘密消息的图像序列后, 需要进行秘密消息的抽取。此时所有秘密消息都存在于载密图像中。神经网络模型是由双方共享, 其中包含图像到码字的映射规则。因此接收方也可以使用相同的深度模型进行秘密消息抽取: 在本文方法中, 接收方收到一系列载密图像后利用神经网络将图像翻译成二进制码段。接收方将所有二进制秘密消息段拼接起来后得到一个完整的二进制秘密消息。最后将整个二进制串秘密消息解密为可读的秘密消息, 完成秘密消息的提取过程。至此, 接收方完成秘密消息的抽取, 整个隐蔽通信的流程结束。

## 4 实验与分析

本文实验涉及多种方案的比较, 而各种方案给出的源代码实现平台不同, 本文为统一比较, 将已开源的方法进行封装, 统一利用 python3.6 在 Windows 环境下实现。其中 DCT+LDA<sup>[35]</sup>方法给出了用 C++在 VS2010 下实现代码; 进行信息隐藏检测的集成分类器在 Matlab2016a 上实现, 本文通过二次封装在 Pycharm 中调用这两种方法的接口。其余涉及的无载体方法皆按照原始论文说明在 Pycharm 环境下复现。本文提出的方法基于 Nvidia Titan 型号 GPU, 采用 pytorch 0.2.03 版本工具箱搭建深度学习模型。实验涉及的网络模型来自于 pytorch 内部封装的标准网络模型, 神经网络的训练参数设置如下: 梯度下降方法选用随机梯度下降, 动量为 0.8, 学习率为 0.001。所有实验图像均来自 ImageNet 2017 test 文件夹。

ImageNet database<sup>[39]</sup>是一个拥有 22000 种类别, 15,000,000 余张图片的数据集, 它由斯坦福大学 Vision Lab 建立。这些图像从网络中收集, 并由 Amazon Mechanical Turk 标记。本实验中图像数据采

用 2017 年发布版本的 test 文件夹中 5491 张图片。

本文方法实验结果主要与 Pixel<sup>[27]</sup>, SIFT+Hash<sup>[28]</sup>, SIFT+BOF<sup>[32]</sup>和 DCT+LDA 对比, 因为 Pixel, SIFT+Hash, SIFT+BOF 等并未给出源码, 我们根据论文介绍复现了这三种方案。DCT+LDA 给出了详细的 c++代码, 本实验通过调用其开源代码进行实验。

### 4.1 容量测试

其中, 无载体信息隐藏的容量取决于码字长度以及每张载体图像的像素数。每张图像表征的码字长度越长或所采用载体图像像素数目越小, 则该方法有越高的容量。本文使用 bpp(bit per pixel)作为指标衡量方法容量。其计算公式为

$$C = \frac{N}{W * H} \quad (3)$$

其中,  $N$  为每张图片可以表示的码字长度,  $W$  为每张载体图像的宽度,  $H$  为每张图像的高度。除了我们选取的几种方法(Pixel、SIFT\_Hash、DCT\_LDA 和 SIFT\_BOF)外, 还收集了李宗翰等<sup>[33]</sup>提供的另外几种方法的容量数据, 制表如下。

表 2 信息隐藏相对容量结果对比表

Table 2 Hide the capacity result table

方法	码字长度	图像尺寸(像素)	相对容量(位/像素)
Pixel	8	512×512	$3.82 \times 10^{-5}$
SIFT_Hash	18	512×512	$6.86 \times 10^{-5}$
DCT_LDA	15	512×512	$5.7 \times 10^{-5}$
SIFT_BOF	8	512×512	$3.82 \times 10^{-5}$
[40]	-	$\geq 512 \times 512$	$1.14 \times 10^{-4}$
[28]	-	512×512	$6.86 \times 10^{-5}$
[15]	-	480×640	$6.51 \times 10^{-4} \sim 2.61 \times 10^{-3}$
[16]	-	800×800	$5.12 \times 10^{-2}$
[19]	-	1024×1024	$1.17 \times 10^{-2} \sim 3.28 \times 10^{-2}$
[14]	-	64×64	$7.33 \times 10^{-3}$
[33]	-	64×64	$2.93 \times 10^{-3}$
Ours	8	5×5	$3.2 \times 10^{-1}$

实验结果如表 2 所示。从表中可以看出, SIFT\_Hash 有着最长的特征序列 18, 但是在现实中准备  $2^{18}$ (即 262144)张图的图像库有较大的时间开销, 考虑到图像生成码字的重复问题, 需要检索的图片将数目远远大于 262144 张, 因此该方法并不适合实际使用。另外虽然我们提出方法的码字长度也为 8, 但是我们所需要的图像尺寸最小仅为 5×5, 从而大大增加了本文方法的容量。因此相较于最近几年提出的几种无载体隐写方法, 我们的方法有着较高的相对容量, 可以用更少的像素传递更多的消息。

图 4 展示了几种方法在传输 1bit,1B,1KB,1MB 信息时所需像素数与秘密消息量的关系。为了方便可视化,对所需像素点数据进行了  $\log_{10}$  归一化。从图中可以形象看出本方法在传递大量消息时所需要的像素数将远远小于其他几种代表方法,也就是说能够以远少于其他方法的像素数传递相同数量的秘密消息。

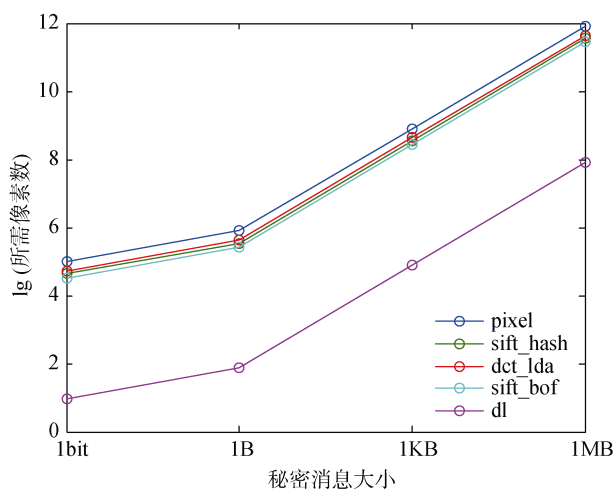


图 4 传递相同数据所需像素数目

Figure 4 The number of pixels required to pass the same data

## 4.2 覆盖率

在传统方法中,不同图片根据某映射规则提取出的  $N$  位长度码字很可能相同,这种频繁的“碰撞”导致实际使用时为了建立一个完备码库,需要收集远多于  $2^N$  张图片来建立候选图像库。除了碰撞以外,也会发生候选图像缺失的问题,例如图像库中缺少对应码字“01000001”的图像,那么需要搜索大量的图像去专门寻找满足该条件的载体图像。由此不仅导致构造图像库的图像要求过高,也造成了巨大的计算资源浪费。本文首先定义覆盖率的概念来衡量候选图像是否存在缺失,其公式定义如下:

$$P = \frac{Num}{M} \times 100\% \quad (4)$$

其中,  $P$  为覆盖率,  $M$  为图像库中载体图像数目,  $Num$  为图像库可产生的不同码字的数目。在相同大小的

图像库上,产生越多的不同码字,则该方法覆盖率越高,那么也就更容易建立一个完备的图像库。本文在 ImageNet2017 test 图像库中随机选取图像构成不同大小的图像库,并测试几种方法覆盖率。

从图 5 中可以看出,本文方法在多次测试中均处于最高的覆盖率—100%。也就是可以为给定的每张图像产生一个唯一的码字,因此本文方法在拥有 256 或 128 张图片时即可形成一个满足要求的图像库而不需要额外再去搜索图片。而其他三种方法随着图像数目增加,其覆盖率也逐渐下降。当给定图像库为 256 张时, DCT+LDA 方法的覆盖率为 43.36%, SIFT+BOF 的覆盖率降低到 61.33%。这说明,此时的图像库中只有一半左右图像可以用作编码,另外一半因产生重复编码而被浪费掉。

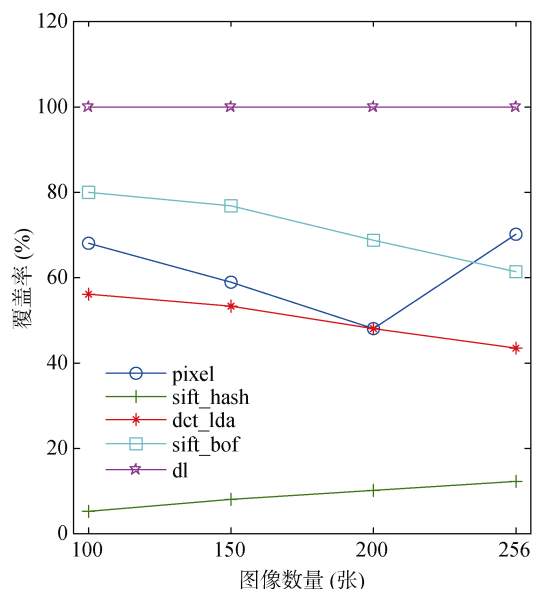


图 5 覆盖率对比

Figure 5 Coverage ratio

此外,重复码字的碰撞现象导致搜索的难度也逐渐增加,表 3 展示了每种方法在不同规模的图像库中获得的可用码字数目。从表 3 中可以看出 LDA+DCT 方法在 100 张图片时可以产生 56 个码字。但是当图像库增长到 200 张时变为 96 个码字,只增长了 40 个,证明了新加入的 100 张图像产生的 100

表 3 去重后码字对比

Table 3 Deduplication code word contrast

图像数量(张)	100	150	200	256
Pixel(8bit)	68	88	96	109
DCT_LDA	56	80	96	111
SIFT_BOF(8bit)	80	115	137	157
Ours	100	150	200	256

个码字中有 60 个与已有的相同产生碰撞。另外在 Pixel 方法中, 图像数量从 100 增加至 150 时, 新增码字数为 20; 图像数量再增加 50 总数达到 200 张时, 新增码字数为 18; 图像数量再增加 56, 总数达到 256 张时, 新增码字数目只有 13。图像每增加 50 张, 产生可利用的码字越来越少。这也证明了传统方法中为了构建一个完备的图像库需要搜索 2 倍甚至几倍于码字数目的图像, 并且随着已知码字越来越多, 搜索的难度将逐渐增大。

该小节实验说明本文方法可以较好解决构建图像库的问题。原因在于本文方法核心是在已经给定的图像库的情况下由神经网络主动去寻找一套高覆盖率的映射方法, 并作为编解码模型。而其他方法是先确定规则, 然后按照规则去搜索图片, 因此不可避免出现大量重复的码字。

### 4.3 抗检测性

基于映射规则的无载体信息隐藏方法不会修改图像内容。因此传送的图片符合真实世界中的数据分布。所以理论上这些图片在网络中传输时也能较好的抵御第三方隐写检测。

为了证明在抗检测性方面, 本文方法优于传统基于修改的方法, 实验首先对 ImageNet 图像库中的图像利用三种经典的隐写算法 nsF5<sup>[41]</sup>, JUNIWARD<sup>[42]</sup>, UERD<sup>[43]</sup>进行 0.1bpp 的消息嵌入, 然后对同一批图像用无载体方法形成载密图像, 由此形成四个带有消息的载密图像集合。

对四个载密图像集合, 每个分成两部分, 一部分用于训练, 另一部分用于测试。我们采用了 ccJRM<sup>[44]</sup>隐写特征提取方法实现隐写图片的特征提取。继而采用 Fridrich 等提出的集成分类器<sup>[45]</sup>进行训练与分类。实验指标用错误率  $Err$  来衡量。

$$Err = \left(1 - \frac{TP + TN}{All}\right) \times 100\% \quad (5)$$

$TP$  表示预测正确的未嵌入消息载体数目,  $TN$  表示预测正确的嵌入消息的载体数目,  $All$  表示所有的样本数量。实验结果如下表 4 所示。从图中可以看出相较于其他方法, 无载体图像有着最好的抗检测性能, 其错误率为 49.64%, 能够使得分类器几乎无法识别网络中传输的载体是否包含了秘密消息。而针对其他几种方法, 分类器均能在一定程度上进行识别。其中最容易被检测的是 nsF5 方法, 其被检测到的错误率为 37.81%。而专门针对 jpeg 格式的图片设计的隐写的方法 JUNIWARD 有着较好的抗检测性, 其错误率为 46.18%, 但是仍然低于无载体 3 个百分点。由此可以看出, 因为无载体方法没有修改载体,

所以无论一张图片是否真的包含了秘密消息, 都能抵御检测方法的攻击, 因此在抗检测性实验中无载体方法的抵抗能力最强。

表 4 检测结果  
Table 4 Detection result

方法	容量(bpp)	错误率(%)
nsF5	0.1	37.81
UERD	0.1	42.04
JUNIWARD	0.1	46.18
Ours	0.1	49.64

进一步分析可以发现, 无载体信息隐藏类方法有着较好抗检测性的根本原因是包含了秘密消息的图像特征分布仍然保持和原始图像的特征一致。载体图像和载密图像特征分布如图 7, 可以看到消息隐藏前后的图像特征分布没有明显变化, 呈现出同分布所以第三方检测工具的分类器无法进行有效区分。而基于修改的信息隐藏方法(nsF5)会在嵌入消息后改变图像的特征分布, 载体图像和载密图像特征分布如图 6 所示以看出, 嵌入消息前后的图像特征分布呈现“双峰”。

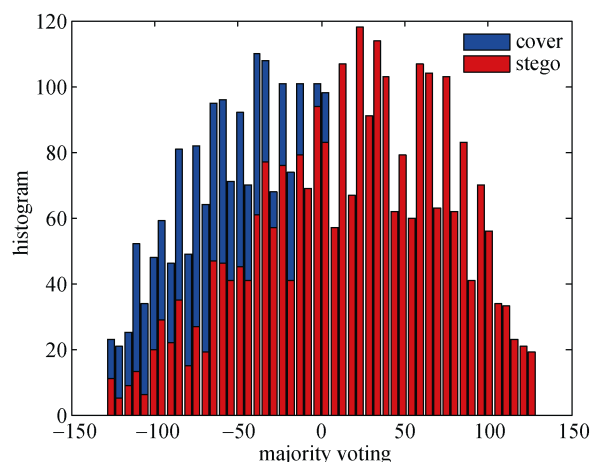


图 6 基于 nsF5 方法嵌入前后特征对比

Figure 6 Nsf5-based feature comparison before and after embedding

### 4.4 抗噪性

鲁棒性与抗噪性是信息隐藏中一个重要的评价指标。其目的在于测试含有秘密消息的图像受到网络中第三方的主动攻击后, 接收方能否将消息正确解码出来。由于攻击者通常会使用一些攻击方法对图像进行变换和破坏, 从而使接收方难以解密出来消息。常见的攻击方式及攻击参数包括以下几种:

a) JPEG 压缩。质量因子包括 10%, 30%, 50%, 70%, 90%。



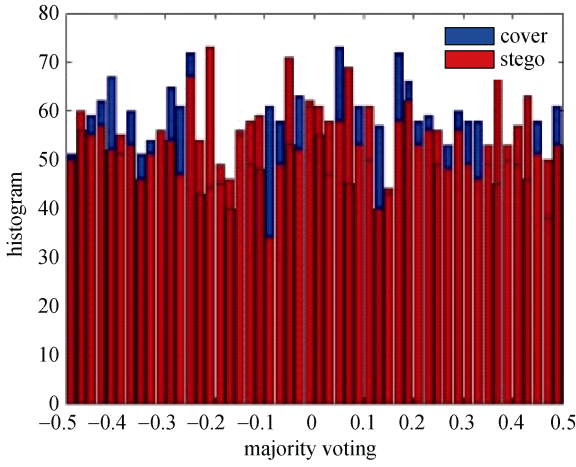


图 7 无载体方法嵌入信息前后特征对比

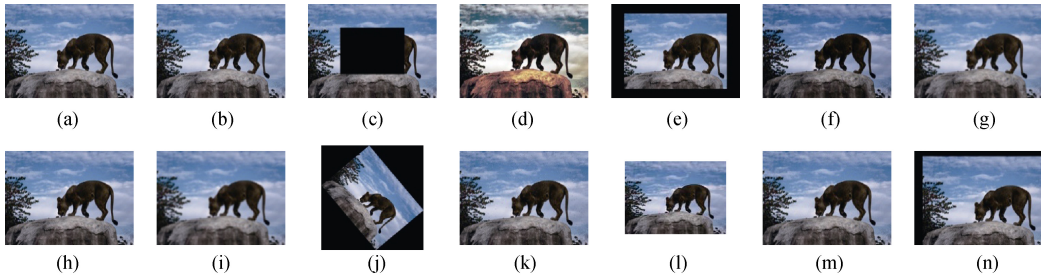
Figure 7 Before and after embedding information with no carrier method

- b) 高斯噪声, 均值 $\mu$ 为 0, 方差 $\sigma$ 包括 0.001, 0.005 和 0.1。
- c) 椒盐噪声, 均值 $\mu$ 为 0, 方差包括 0.001, 0.005 和 0.1。
- d) 散斑噪声, 均值 $\mu$ 为 0, 方差包括 0.01, 0.05 和 0.1。
- e) 均值滤波, 均值窗口包括  $3 \times 3$ ,  $5 \times 5$  和  $7 \times 7$
- f) 中值滤波器, 滤波窗口大小为  $3 \times 3$ ,  $5 \times 5$  和  $7 \times 7$ 。
- g) 高斯低通滤波器, 滤波窗口为  $3 \times 3$ ,  $5 \times 5$  和  $7 \times 7$ 。

- h) 中心裁剪, 裁剪率包括 20%, 50%。
- i) 边缘裁剪, 裁剪率包括 10% 和 20%。
- j) 旋转, 旋转角度包括  $10^\circ$ ,  $30^\circ$  和  $50^\circ$ 。
- k) 平移, 因为选定的图像平移包括 (8, 5), (16, 10) 和 (32, 20)。
- l) 缩放, 缩放率包括 0.3, 0.5, 0.75, 1.5 和 3。
- m) 颜色直方图均衡化。
- n) 伽马矫正。

为了形象展示图像经过各种攻击算法攻击之后的变化, 在图 8 中展示了一张图像被施加各种算法攻击之后的结果。

本实验在图像库中随机选取若干张图片, 构成鲁棒性测试图像库, 用不同的无载体隐写算法进行多次实验测试, 获得载密图像库。随后对载密图像库进行多种类别、多种参数的攻击来模仿网络通信中的噪声或第三方恶意攻击, 获得被攻击图像库。然后通过对比载密图像库和被攻击图像库中对应图片的解码消息是否一致, 测定无载体隐写算法的鲁棒性。由于隐蔽通信中更加重视如何将每个语义字符正确解析出来, 因此误码率并不是很好的衡量指标, 例如字符 ‘a’ 对应 ASCII 码为 01100001, 如果发生一位错误变为 01101001, 虽然误码率仅为 12.5%, 但是随之解析出来的字符为 ‘i’, 导致语义上产生了错误。因此实验中采用 CA 表示字符级别解码的正确率(Character-level decoding Accuracy), CA 计算公式如下:



(a) original, (b) jpeg compress50, (c) CenterCropping(0.5), (d) ColorHist, (e) EdgeCropping(0.2), (f) GammaCorrection(0.8), (g) GaussFilter(7x7), (h) GaussNoise(0.1), (i) MeanFilter(7x7), (j) Rotation(50), (k) Salt&Pepper(0.005), (l) Scaling(0.75), (m) SPeckleNoise(0.05), (n) Translation

图 8 部分受攻击图像

Figure 8 Part of attacked images

$$CA = \left( 1 - \frac{\sum_{sg=1}^{sg=m} f(bs, bs')}{m} \right) \times 100\% \quad (6)$$

公式中  $m$  表示所有参与解码图片的数目,  $sg$  为图片的序号,  $bs$  表示发送方嵌入的消息,  $bs'$  表示接收方解码出的消息。如果  $bs$  与  $bs'$  相等, 则  $f(bs, bs')$  返回 1, 否

则返回 0。

首先我们对比几种经典神经网络模型的表现。我们采用了 Resnet18<sup>[46]</sup>, Vgg11<sup>[47]</sup>, Alexnet<sup>[48]</sup> 三种神经网络模型, 并在图像库中进行训练, 各自搜索一套最佳映射算法, 直到模型能够将全部数据正确分类。然后取攻击后的图像送入网络进行测试。实验结果如表 5 所示。从表中我们可以看出, 不同的网络结构在抵抗大类噪声攻击方面表现接近, 例如三种方法抵抗

表 5 不同深度模型的解码正确率比较

Table 5 The decoding accuracy of different depth models is compared (%)

Attack Method	vgg11	resnet	alexnet
CenterCropping(0.2)	14.12	0.39	17.25
CenterCropping(0.5)	71.76	61.18	72.94
ColorHist	51.37	76.47	66.27
Compressing(10)	0.00	0.00	0.39
Compressing(30)	0.00	0.00	0.39
Compressing(50)	0.00	0.00	0.39
Compressing(70)	0.00	0.00	0.39
Compressing(90)	0.00	0.00	0.39
EdgeCropping(0.1)	1.96	1.57	0.39
EdgeCropping(0.2)	47.45	3.14	8.24
GammaCorrection(0.8)	0.78	1.18	24.71
GaussFilter(3×3)	0.39	1.96	1.18
GaussFilter(5×5)	1.18	11.37	5.88
GaussFilter(7×7)	6.27	27.06	14.12
GaussNoise(0.001)	0.00	0.00	0.00
GaussNoise(0.005)	0.00	0.00	0.00
GaussNoise(0.1)	0.00	0.00	0.00
MeanFilter(3×3)	0.39	4.31	3.14
MeanFilter(5×5)	7.84	29.80	16.08
MeanFilter(7×7)	23.92	49.80	38.04
MedianFiltering(3×3)	0.39	0.00	0.39
MedianFiltering(5×5)	0.78	12.94	5.88
MedianFiltering(7×7)	3.53	27.45	11.76
original	0.00	0.00	0.00
Rotation(10)	49.02	2.35	22.35
Rotation(30)	92.55	38.82	90.98
Rotation(50)	98.04	65.88	95.29
Salt&Pepper(0.001)	0.00	0.39	0.00
Salt&Pepper(0.005)	0.00	0.78	0.39
Salt&Pepper(0.1)	7.45	68.63	12.55
Scaling(0.3)	21.96	40.78	35.29
Scaling(0.5)	6.67	18.43	13.73
Scaling(0.75)	1.57	3.14	3.92
Scaling(1.5)	0.39	0.00	0.39
Scaling(3.0)	0.00	0.00	0.00
SPeckleNoise(0.01)	0.00	0.00	0.00
SPeckleNoise(0.05)	0.00	0.39	0.00
SPeckleNoise(0.1)	0.00	1.57	0.00
Translation(16_10)	0.39	0.78	1.57
Translation(32_20)	14.12	1.57	15.29
Translation(8_5)	0.00	0.39	0.78
average	12.79	13.48	14.17

jpeg 压缩性能较好, 在被攻击后基本不受 jpeg 影响, 均能正确解码出信息; 但是所有方法受旋转攻击影响较大, 三种方法均会产生一半以上的错误率。但是

对于某些种类攻击方法, 不同攻击参数下也有模型会存在一些相对较好的表现, 例如受到中值滤波攻击时, Vgg11 表现出较好的抗中值滤波攻击性能,

Resnet18 在抵抗中心裁剪 20%时表现远胜于另外两种模型。按照平均解码错误率来看, Vgg11 是表现最优的方法。

随后, 我们进行了深度学习方法与现有传统方法鲁棒性的实验对比, 结果呈现在表 6 中。从表中结果来看, 我们的方法相对传统方法抗噪声攻击的表

表 6 不同无载体方法 CA  
Table 6 Different carrier free methods CA

(%)

CA	SIFT_Hash	SIFG+BOW+HASH	Pixel	DCT+LDA	vgg11
CenterCropping(0.2)	87.11	99.22	54.30	31.64	<b>14.12</b>
CenterCropping(0.5)	99.61	99.61	92.97	82.03	<b>71.76</b>
ColorHist	99.22	98.83	32.03	<b>27.34</b>	51.37
Compressing(10)	84.77	99.61	0.78	1.17	<b>0.00</b>
Compressing(30)	84.77	99.61	0.78	1.17	<b>0.00</b>
Compressing(50)	84.77	99.61	0.78	1.17	<b>0.00</b>
Compressing(70)	84.77	99.61	0.78	1.17	<b>0.00</b>
Compressing(90)	84.77	99.61	0.78	1.17	<b>0.00</b>
EdgeCropping(0.1)	98.44	98.83	73.05	67.19	<b>1.96</b>
EdgeCropping(0.2)	99.22	99.61	89.45	85.55	<b>47.45</b>
GammaCorrection(0.8)	93.36	99.61	10.94	7.81	<b>0.78</b>
GaussFilter(3×3)	99.61	100.00	<b>0.00</b>	0.00	0.39
GaussFilter(5×5)	99.61	100.00	<b>0.00</b>	0.00	1.18
GaussFilter(7×7)	99.22	100.00	<b>0.00</b>	0.00	6.27
GaussNoise(0.001)	73.83	99.61	0.00	0.78	<b>0.00</b>
GaussNoise(0.005)	73.83	99.61	0.00	0.78	<b>0.00</b>
GaussNoise(0.1)	73.83	99.22	0.00	0.78	<b>0.00</b>
MeanFilter(3×3)	99.61	100.00	<b>0.00</b>	0.00	0.39
MeanFilter(5×5)	99.22	100.00	<b>0.00</b>	0.00	7.84
MeanFilter(7×7)	99.61	100.00	<b>0.00</b>	1.17	23.92
MedianFiltering(3×3)	98.44	99.61	1.56	3.13	<b>0.39</b>
MedianFiltering(5×5)	99.22	98.83	4.69	4.69	<b>0.78</b>
MedianFiltering(7×7)	99.61	99.22	5.08	4.69	<b>3.53</b>
Original	0.00	0.00	0.00	0.00	<b>0.00</b>
Rotation(10)	99.61	99.22	89.06	86.72	<b>49.02</b>
Rotation(30)	100.00	100.00	94.53	96.48	<b>92.55</b>
Rotation(50)	100.00	100.00	97.27	<b>97.27</b>	98.04
Salt&Pepper(0.001)	83.59	100.00	0.39	0.78	<b>0.00</b>
Salt&Pepper(0.005)	95.70	99.61	0.39	1.17	<b>0.00</b>
Salt&Pepper(0.1)	100.00	99.61	4.30	<b>7.03</b>	7.45
Scaling(0.3)	100.00	99.22	<b>0.78</b>	1.17	21.96
Scaling(0.5)	99.22	98.83	<b>0.39</b>	0.78	6.67
Scaling(0.75)	97.27	99.61	<b>0.00</b>	0.78	1.57
Scaling(1.5)	95.70	98.83	0.00	0.78	0.39
Scaling(3.0)	94.14	98.44	0.39	0.78	<b>0.00</b>
SPeckleNoise(0.01)	83.98	100.00	0.39	0.39	<b>0.00</b>
SPeckleNoise(0.05)	97.66	99.61	1.17	0.39	<b>0.00</b>
SPeckleNoise(0.1)	99.22	100.00	1.95	1.56	<b>0.00</b>
Translation(16_10)	100.00	99.22	56.64	58.98	<b>0.39</b>
Translation(32_20)	100.00	99.61	79.69	82.03	<b>14.12</b>
Translation(8_5)	100.00	99.61	35.94	37.50	<b>0.00</b>
Average	91.77	97.10	20.27	19.46	<b>12.79</b>

现具有较明显的优势。在传统的四种方法中, 表现较好的是 Pixel 和 DCT+LDA, 而另外两种采用 Hash 算法的模型则表现较差, 错误率在 80% 以上。这是由于采用了 Hash 函数而无法避免的结果: 只要输入图像加入了少量噪声而有微小变化那么依赖 Hash 算法的输出将会有巨大的变化, 从而导致无法正确解码得到原始消息。在实验结果的对比中, 可以看到在 jpeg 压缩, 高斯噪声, 椒盐噪声等攻击下, 深度学习的方法可以达到最优水平, 甚至在 jpeg 攻击下解码错误率为 0; 在传统方法表现较差的几种攻击中本文方法仍然有较好的表现, 例如中心裁剪攻击, 边缘裁剪攻击, 旋转攻击等; 最后从平均错误率综合来看, 本文方法有着最低的平均错误率。说明本文方法在鲁棒性方面优于传统人工设计映射规则的无载体信息隐藏方法。

通过表 5 分析我们发现使用不同的网络可以在不同抗噪性上有所差异, 因此后续通过继续调整、组合不同网络的结构, 我们可以综合各个方法的优势, 尝试搭建更加鲁棒的深度模型。

#### 4.5 方案优势

从上述介绍可以看出我们的方案有以下几个优势: 1) 依据该规则进行的隐蔽通信能抵抗隐写检测工具。在网络中进行信息隐藏首先就是要保证通信的安全性, 一旦被检测出来存在信息隐藏的痕迹, 即视为隐蔽通信系统的失败。我们的方案不修改载体, 从理论上可证明无法被检测; 2) 较高的覆盖率, 可以避免大量图像搜索, 降低隐蔽通信的计算开销。在构建码表部分, 我们充分利用每一张图片的编码潜力, 可以做到尽量少的图像浪费; 3) 能够快速找到一组合适的映射规则, 支持快速生成以及方便的共享, 降低通信初始的代价。通过使用 GPU 来训练模型, 可以快速训练出符合要求的模型; 4) 较高的容量, 由于隐蔽通信每次通信的代价较大, 因此要求通信的方法要尽量提高容量, 以减少发送次数和每次发送的载体数量, 从而进一步降低被检测隐蔽通信的可能性; 5) 鲁棒性高, 网络传输中经常存在针对图像的压缩、重建等操作, 甚至有第三方攻击者对图像的恶意盲攻击, 而我们方法在受到攻击的时候仍然能够解码出准确率较高的消息; 6) 映射规则隐蔽性好, 本文方法的映射规则保存在深度学习模型中, 可以通过拷贝、硬件预置等方式实现共享, 即使被截获也很难利用该模型的矩阵参数恢复出映射规则, 因此该方案有较好的隐蔽性; 7) 较高的行为隐蔽性, 本方案中传递的图像可以做到语义相关, 来防止针对行为模式的侧信道攻击。

另外, 本文方法支持映射规则多样性, 即发送方可以对同一个图像库按照不同码字的分配方案来指定多种映射规则, 达到不同映射规则对应不同隐蔽通信接收方的目的。因为对神经网络来说, 在分类问题上将猫狗图像标记为 0 和 1 或者 1 和 0, 其实没有不同, 但是从编码角度看这是两种不同的映射规则。因此对同一组图像库执行不同的码字分配规则, 经过训练就可以得到多个映射规则的模型, 可以有效保证秘密消息对不同接收方的安全隔离。

## 5 结论

本文提出了一种基于深度学习的无载体映射关系自动搜索方案。其主要利用神经网络进行自动学习的方法来找到一个拥有最高覆盖率的映射方案。实验结果表明, 本文提出的方案在容量上比目前大多数传统无载体信息隐藏方法更高。另外本文方法不需要人工设计复杂特征, 在抗检测性, 鲁棒性方面超过了传统人工设计的方案。本文方法也可以解决之前提出的基于关系映射的搜索式无载体方法必须携带一个大型图像库的问题, 使得无载体通信可以进一步迈向实用。

## 6 问题与展望

首先, 虽然本文方案的相对容量在无载体方法中处于较高水平, 但是为了推动无载体隐写方法得到更广泛的应用, 未来需要进一步提高每幅图像所能隐藏的相对容量。因为在实验中我们发现码字长度每增加 1 位, 对应图片集合规模与训练时间将会成指数级增加, 所以在现有工作中我们仅使用了几百张图片进行 8 位长度码字的映射。我们考虑后续通过利用编码多样性构建基于多维度的编码方案, 在保持图片集合规模基本不变的前提下, 丰富单个码字对应的编码内容实现容量的进一步提高。

另一方面, 我们看到不同网络模型结构, 在抵抗噪声攻击时表现各有优劣。例如 Vgg11 网络可以较好地抵抗过滤器攻击, 而 Resnet18 在图像平移攻击方面表现优异。因此未来可以通过融合多种神经网络结构, 结合每种结构的优势, 使得网络进一步提高鲁棒性能。

**致 谢** 在此向本文成文中给予指导的老师、提供帮助的同学和给本文提出建议的评审专家表示感谢。

## 参考文献

- [1] C E Shannon. Communication theory of secrecy systems[J]. Bell



- system technical journal, 1949, 28(4): 656-715.
- [2] Simmons G J. The Prisoners' Problem and the Subliminal Channel[M]. *Advances in Cryptology*. Boston, MA: Springer US, 1984: 51-67.
  - [3] I Goodfellow, Pouget-Abadie, J, Mirza M, et al. Generative adversarial nets[C]. *Advances in neural information processing systems*. 2014: 2672-2680.
  - [4] Zhang X P, Qian Z X, Li S. Prospect of Digital Steganography Research[J]. *Journal of Applied Sciences*, 2016, 34(5):475-489.  
(张新鹏, 钱振兴, 李晟. 信息隐藏研究展望[J]. *应用科学学报*, 2016, 34(5):475-489.)
  - [5] Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. *Neural Computation*, 2006, 18(7):1527-1554.
  - [6] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based Learning Applied to Document Recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.
  - [7] Hinton G. Deep Belief Networks[J]. *Scholarpedia*, 2009, 4(5):5947.
  - [8] Vincent P, Larochelle H, Lajoie I, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion[J]. *Journal of Machine Learning Research*, 2010, 11:3371-3408.
  - [9] Hochreiter S, Schmidhuber J. Long Short-term Memory[J]. *Neural Computation*, 1997, 9(8):1735-1780.
  - [10] Liu M M, Zhang M Q, Liu J, et al. Generative Steganography Based on GANs[M]. *Cloud Computing and Security*. Cham: Springer International Publishing, 2018: 537-549.
  - [11] X Duan, H Song, C Qin, et al. Coverless steganography for digital images based on a generative model[J]. *Computers, Materials & Continua*, 2018, 55(3): 483-493.
  - [12] Chu, C., Zhmoginov, A., Sandler, M. CycleGAN, a master of steganography. arXiv preprint arXiv:1712.02950, 2017.
  - [13] Liu J, Zhou T P, Zhang Z, et al. Digital Cardan Grille: A Modern Approach for Information Hiding[C]. *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*. 2018: 441-446.
  - [14] Hu D H, Wang L, Jiang W J, et al. A Novel Image Steganography Method via Deep Convolutional Generative Adversarial Networks[J]. *IEEE Access*, 2018, 6:38303-38314.
  - [15] Otori H, Kuriyama S. Texture Synthesis for Mobile Data Communications[J]. *IEEE Computer Graphics and Applications*, 2009, 29(6):74-81.
  - [16] Xu J Y, Mao X Y, Jin X G, et al. Hidden Message in a Deformation-based Texture[J]. *The Visual Computer*, 2015, 31(12): 1653-1669.
  - [17] Lu S F, Jaffer A, Jin X G, et al. Mathematical Marbling[J]. *IEEE Computer Graphics and Applications*, 2012, 32(6):26-35.
  - [18] Qian Z X, Pan L, Li S, et al. Steganography by Constructing Marbling Texture[M]. *Cloud Computing and Security*. Cham: Springer International Publishing, 2018: 428-439.
  - [19] Wu K C, Chung-Ming W. Steganography Using Reversible Texture Synthesis[J]. *IEEE Transactions on Image Processing*, 2015, 24(1):130-139.
  - [20] QL Zhou, YB Qiu, L Li, et al. Steganography using reversible texture synthesis based on seeded region growing and LSB[J]. *Comput. Mater. Continua*, 2018, 55(1): 151-163.
  - [21] Qin Z C, Li M, Wu B. Robust Steganography via Patch-Based Texture Synthesis[M]. *Communications in Computer and Information Science*. Singapore: Springer Singapore, 2018: 429-439.
  - [22] Wei W Y, A C, Wang L Z, et al. A Texture Synthesis Steganography Scheme Based on Super-Pixel Structure and SVM[J]. *Intelligent Information Processing IX*, 2018: 375-383.
  - [23] Qian Z X, Zhou H, Zhang W M, et al. Robust Steganography Using Texture Synthesis[M]. *Advances in Intelligent Information Hiding and Multimedia Signal Processing*. Cham: Springer International Publishing, 2016: 25-33.
  - [24] Lee W K, Ong S, Wong K, et al. A Novel Coverless Information Hiding Technique Using Pattern Image Synthesis[C]. *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 2018: 1122-1127.
  - [25] Zhang X, Peng F, Lin Z X, et al. A Coverless Image Information Hiding Algorithm Based on Fractal Theory[J]. *International Journal of Bifurcation and Chaos*, 2020, 30(4):2050062.
  - [26] Fridrich J. Steganography in Digital Media[M]. Cambridge: Cambridge University Press, 2009.
  - [27] Zhou Z L, Sun H Y, Harit R, et al. Coverless Image Steganography without Embedding[M]. *Cloud Computing and Security*. Cham: Springer International Publishing, 2015: 123-132.
  - [28] Y Cao, Z Zhou, X Sun, et al. Coverless information hiding based on the molecular structure images of material[J]. *Computers, Materials & Continua*, 2018, 54(2): 197-207.
  - [29] Sivic J, Zisserman A. Efficient Visual Search of Videos Cast as Text Retrieval[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(4): 591-606.
  - [30] Lowe D G. Object recognition from local scale-invariant features[C]. *iccv*. 1999, 99(2): 1150-1157.
  - [31] Zheng S L, Wang L, Ling B H, et al. Coverless Information Hiding Based on Robust Image Hashing[M]. *Intelligent Computing Methodologies*. Cham: Springer International Publishing, 2017: 536-547.
  - [32] Zhou Z L, Mu Y, Wu Q M J. Coverless Image Steganography Using Partial-duplicate Image Retrieval[J]. *Soft Computing*, 2019, 23(13):4927-4938.
  - [33] Li Z H, Liu J, Ke Y, et al. Cover Selection Steganography Scheme Based on Image-to-Image Translation[J]. *Journal of Applied Sciences*, 2019, 37(5): 733-743.

(李宗翰, 刘佳, 柯彦, 等. 基于图像翻译的载体选择式图像隐写方案[J]. *应用科学学报*, 2019, 37(5):733-743.)

- [34] Zou L M, Sun J D, Gao M, et al. A Novel Coverless Information Hiding Method Based on the Average Pixel Value of the Sub-images[J]. *Multimedia Tools and Applications*, 2019, 78(7):7965-7980.
- [35] Zhang X, Peng F, Long M. Robust Coverless Image Steganography Based on DCT and LDA Topic Classification[J]. *IEEE Transactions on Multimedia*, 2018, 20(12):3223-3238.
- [36] Wu J B, Liu Y W, Dai Z W, et al. A Coverless Information Hiding Algorithm Based on Grayscale Gradient Co-occurrence Matrix[J]. *IETE Technical Review*, 2018, 35(sup1):23-33.
- [37] Y Cao, Z Zhou, CN Yang, et al. Dynamic content selection framework applied to coverless information hiding[J]. *Journal of Internet Technology*, 2018, 19(4): 1179-1186.
- [38] Zhou Z L, Cao Y, Wang M M, et al. Faster-RCNN Based Robust Coverless Information Hiding System in Cloud Environment[J]. *IEEE Access*, 2019, 7:179891-179897.
- [39] J Deng, W Dong, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]. 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.
- [40] Zhou Z L, Cao Y, Sun X M. Coverless Information Hiding Based on Bag-of-Words Model of Image[J]. *Journal of Applied Sciences*, 2016, 34(5): 527-536.
- (周志立, 曹巍, 孙星明. 基于图像 Bag-of-Words 模型的无载体信息隐藏[J]. *应用科学学报*, 2016, 34(5):527-536.)
- [41] J Fridrich J, Pevný T, Kodovský J. Statistically Undetectable Jpeg Steganography: Dead Ends Challenges, and Opportunities[C]. *Proceedings of the 9th Workshop on Multimedia & Security - MM&Sec '07*. 2007:3-14.
- [42] V. Holub, J. Fridrich, T. Denemark, Universal Distortion Function for Steganography in an Arbitrary Domain[J]. *EURASIP Journal on Information Security*, 2014(1):25-36.
- [43] Guo L J, Ni J Q, Su W K, et al. Using Statistical Image Model for JPEG Steganography: Uniform Embedding Revisited[J]. *IEEE Transactions on Information Forensics and Security*, 2015, 10(12):2669-2680.
- [44] J. Kodovsky, J. Fridrich. Steganalysis of JPEG Images Using Rich Models[C]. *SPIE*, 2012:125-136.
- [45] J. Kodovský, J. Fridrich, Steganalysis in high dimensions: fusing classifiers built on random subspaces[C]. *SPIE*, 2011: 365-372.
- [46] K He, X Zhang, S Ren, et al. Deep residual learning for image recognition[C]. *IEEE conference on computer vision and pattern recognition*, 2016: 770-778.
- [47] Simonyan, Karen, Zisserman, et al. Very Deep Convolutional Networks for Large-Scale Image Recognition 2014:arXiv:1409.1556.
- [48] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. *Communications of the ACM*, 2017, 60(6):84-90.



**王亚宁** 于 2018 年在河北工业大学软件工程专业获得学士学位。现在中国科学院信息工程研究所计算机技术专业攻读硕士学位。研究领域为多媒体内容安全、深度学习与信息隐藏。Email: wangyaning@iie.ac.cn



**吴彬** 于 2009 年在中国科学院软件技术研究所获得博士学位, 现任中国科学院信息工程研究所副研究员。研究领域为隐蔽通信、区块链技术以及网络协议分析等。Email: wubin@iie.ac.cn