DOI:10.19651/j.cnki.emt.2415756

以神经网络模型为载体的鲁棒隐写方法*

杨彤彤 杨紫云 王子驰
(上海大学通信与信息工程学院 上海 200444)

摘 要:神经网络已广泛应用于各个领域,神经网络模型隐写是近年来学术界新兴的研究方向。嵌入容量与鲁棒性是神经网络模型隐写的重要指标,但难以同时兼顾。为此,本文提出了一种以神经网络模型为载体的鲁棒模型隐写方法。不明显降低模型原始任务性能的情况下,发送者在训练过程中将秘密信息嵌入到神经网络中,而不是在神经网络训练完成后通过修改网络参数嵌入。接收者使用解码网络提取秘密信息,解码网络的参数使用唯一的嵌入密钥生成,因此无需秘密地向接收者传送解码网络。此外,本文还引入了 RS 码,提高数据提取的鲁棒性。实验结果表明,所提出的模型隐写方法将嵌入容量增大了 66.6%的同时增强了鲁棒性。

关键词:隐写;神经网络模型;鲁棒性;RS 码;嵌入容量
中图分类号:TN918.91 **文献标识码**:A **国家标准学科分类代码**:520.2060

Robust steganography for neural network models

Yang Tongtong Yang Ziyun Wang Zichi
(School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: Neural networks have been extensively utilized in various fields, steganography for neural network is a research emerging direction in academia in recent years. Embedding capacity and robustness are important indicators for steganography. But balancing embedding capacity and robustness is challenging. This paper proposes a robust steganography for neural network models. Embedding secret data into neural network without visibly reducing the performance of the original task. This is achieved by embedding secret data during the training process instead of modifying the network parameters after training. Receivers can obtain the secret data from data decoding networks, the parameters of data decoding networks are generated using the embedding keys. In this way, it is unnecessary to transmit the decoding networks secretly. Additionally, introducing reed-solomon codes to improve data extraction robustness. Experimental results reveal that the robust steganography for neural models improves robustness while maintaining superior embedding capacity.

Keywords: steganography; neural network models; robustness; reed-solomon codes; embedding capacity

0 引 言

隐写术是一种隐蔽通信技术[1],可以将秘密信息隐藏于图像[2]、音频[3]、视频[4]等公开媒体中,进行存储或传输。隐写术在不引起注意的情况下通过轻微修改,将秘密信息嵌入到原始载体数据中[5-7],并在公共信道上秘密传输而不引起怀疑,实现信息的隐蔽传输和储存。Wu 等[8]提出一种基于生成对抗网络(generative adversarial network, GAN)的卡登格子方案,在图像的破损区域嵌入信息。Hu 等[9]提出基于深度卷积对抗网络(deep convolutional generative adversarial networks, DCGAN)图像隐写方法,可以在发送者把秘密信息隐藏在噪声数据中,接收者可提取出所隐藏的秘密信息。文献[10]通过训练神经网络使得多个秘密图像隐藏在载体图像中,并且选择载体图像的 RGB 通道作为嵌入空间。随着神经网络的不断发展,隐写的功能也变得更加强大和多样化。目前,神经网络已被广泛应用于各个领域,如目标跟踪[11]、自然语言处理[12]、计算机视觉任务[13]等。神经网络的广泛使用也使得神经网络模型隐写具有巨大潜力。神经网络模型隐写以神经网络模型作为载体,发送者将秘密信息嵌入到模型中,并通过公共信道传送,如 GitHub 模型池、Google、drive 等。接收者可从公共信道下载含密的隐写网络并提取秘密信息。普通用户可下载该隐写网络执行常规的机器学习任务。发送者将

收稿日期:2024-04-02
* 基金项目:国家自然科学基金(62376148)、上海市教育发展基金会和上海市教育委员会“晨光计划”项目(22CGA46)资助

隐写网络伪装成使用密钥执行常规机器学习任务的正常模型,并在训练过程中嵌入秘密信息。携带秘密信息的隐写网络通过公共信道传送至接收者,接收者可使用正确密钥提取相应部分的秘密信息,无法确定其他部分的秘密信息。

网络模型作为载体在人工智能领域取得了巨大的成功,在通过数字媒体秘密传输秘密信息的同时,不会导致载体媒体的严重失真[14]。有研究者通过网络模型为载体实现信息隐藏,文献[15]对原始图像集的概率密度函数进行建模,将特定位置的秘密图像嵌入到原始图像集的概率分布中,从而将图像隐藏在神经网络中。文献[16]中使用后门技术训练网络模型,通过触发器故意输出特定数据而对主要任务没有明显的影响。文献[17]提出多个发送者同时向一个接收者发送不同秘密数据的多源数据隐藏方案,提出基于提取器匹配的神经网络隐写术。Zhu 等[18]提出了一种通过神经网络学习来编码大量有用信息的方案,使用不可见的扰动来执行数据隐藏任务。然而大容量数据隐藏方案往往会导致封面图像颜色、轮廓等失真。因此嵌入容量具有重要意义。Goodfellow 等[19]提出的生成对抗网络 GAN 实现两个网络互相对抗并交替训练,但在嵌入容量方面还存在不足。总体而言,上述提出的方法嵌入容量较小,鲁棒性不佳。为此本文提出了以神经网络模型为载体的鲁棒隐写方法,改善网络模型嵌入容量和鲁棒性。

在使用模型隐写术实现隐蔽通信中,发送者通过公共信道将含密的隐写网络传输给接收者。实际公共信道有损,本文提出了以神经网络模型为载体的鲁棒隐写方法。秘密信息在神经网络训练过程中完成嵌入和提取,而不是训练结束后修改网络参数嵌入。发送者在对原始模型影响很小的情况下将秘密数据嵌入到神经网络模型中。接收者通过使用正确的嵌入密钥获取相应的秘密信息,对于秘密信息的其他部分,接收者无法确定其存在,因此无法提取其他部分的秘密信息。提取到的秘密信息使用 RS 码(reed-solomon codes,RS 码)进行纠错。结果显示,与文献[20]相比添加噪声情况下秘密信息提取误差降低,网络模型具有较好的抗鲁棒性。隐写模型嵌入容量增大。实验结果说明提出的方法进一步提升了信息隐藏的嵌入容量和鲁棒性。

本文的主要贡献如下:

1)提出了一种以神经网络模型为载体的鲁棒隐写方法。在训练过程中将秘密信息嵌入网络,通过公共信道传输含密的隐写网络,解码网络参数通过嵌入密钥确定,无需秘密存储和传输解码网络。提高了网络模型的实用性。

2)提出了一种可制定的信息隐藏方法。通过神经网络向接收者传递秘密信息时,接收者使用正确密钥可以提取相应部分秘密信息,无法识别其他部分的存在。

信道接收到的秘密信息使用 RS 码进行纠错。RS 码使用冗余信息扩展来恢复原始消息,降低提取错误率。提高信息隐藏的鲁棒性。

1 神经网络隐藏方法

1.1 整体框架

1) 基本框架

本文提出了一种以神经网络模型为载体的鲁棒隐写方法。在公共信道中完成神经网络模型隐写,并提升网络模型的嵌入容量和鲁棒性。如图 1 所示,发送者使用数据隐藏密钥 $\{\boldsymbol{K}_1,\boldsymbol{K}_2,\cdots,\boldsymbol{K}_n\}$,向训练过程中神经网络的预测向量 $\overline{\boldsymbol{u}}_{j|i}$ 嵌入元素的秘密信息 $\{m_1,m_2,\cdots,m_n\}$。接收者使用正确的嵌入密钥生成解码网络参数提取秘密信息 $\{\boldsymbol{M}_1,\boldsymbol{M}_2,\cdots,\boldsymbol{M}_n\}$,不需要秘密地储存和传输解码网络。为验证提出方案鲁棒性,在嵌入网络中添加具有高斯分布和均匀分布的噪声模拟实际有损信道,接收到的秘密信息使用 RS 码进行纠错,最终获得原始信息 $\boldsymbol{r}'_{I_i}$。

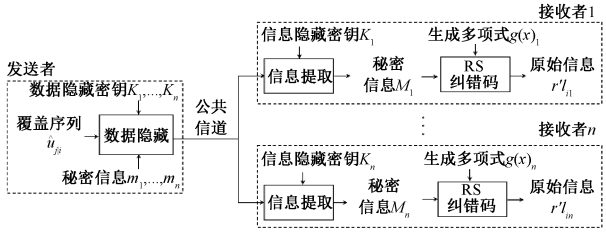


图 1 神经网络模型为载体的鲁棒隐写框架

使用神经网络实现隐蔽通信,秘密信息序列通过网络参数的覆盖序列嵌入。一个接收者的嵌入函数表示为:

$$\boldsymbol{P}_M = f(P_r,k) \tag{1}$$

其中,P_r 是被嵌入的信息,k 是接收者拥有的密钥。给定一个秘密序列 $\boldsymbol{M}=[m_1,m_2,\cdots,m_i,\cdots,m_n]^{\mathrm{T}}\in[0,1]^{\mathrm{T}}$。输出 $\boldsymbol{P}_M=[p_m(1),p_m(2),\cdots,p_m(t)]^{\mathrm{T}}\in[0,1]^{\mathrm{T}}$,通过添加额外的正则化项 $\boldsymbol{P}_M$ 使接近 $\boldsymbol{M}$。同样的,n 个接收器的情况下的嵌入函数表示为:

$$\{\boldsymbol{P}_{M1},\boldsymbol{P}_{M2},\cdots,\boldsymbol{P}_{Mn}\} = f(P_r,K_1,K_2,\cdots,K_n) \tag{2}$$

其中,$\boldsymbol{K}=\{K_1,K_2,\cdots,K_i,\cdots,K_n\}\in[0,1]$ 是 n 个接收器的嵌入密钥。发送者输出 $\{\boldsymbol{P}_{M1},\boldsymbol{P}_{M2},\cdots,\boldsymbol{P}_{Mn}\}$ 应接近于 $\{\boldsymbol{M}_1,\boldsymbol{M}_2,\cdots,\boldsymbol{M}_n\}$,其中 $\boldsymbol{M}_r=[m_r(1),m_r(2),\cdots,m_r(n)]^{\mathrm{T}}\in[0,1]^{\mathrm{T}}$,$\boldsymbol{P}_M=[p_{m_r}(1),p_{m_r}(2),\cdots,p_{m_r}(t)]^{\mathrm{T}}\in[0,1]^{\mathrm{T}}$,$r\in\{1,2,\cdots,n\}$。接收到的数据使用 RS 码进行纠错。RS 码将接收器输出原始消息 $\boldsymbol{M}$ 展开为一维向量,则原始消息多项式为:

$$\boldsymbol{M}(x)=\sum_{i=0}^{k-1}\boldsymbol{M}x^i \tag{3}$$

在有限域 F 中,初始元素为 α,域中数值表示为 $\{0,\alpha^0,\alpha^1,\alpha^2,\cdots,\alpha^{|F|-2}\}$。编码器根据式(4)生成多项式附加在原始信息后。

$$g(x)=\prod_{i=0}^{m-1}x-\alpha^i \tag{4}$$

其中,m 为附加到消息后纠错码的长度,$1\leqslant m\leqslant$

$|F|-k$ 使用式(5)完成编码后将码字发出。

$$\boldsymbol{s}(\boldsymbol{x})=\boldsymbol{M}(\boldsymbol{x})x^{m}-[\boldsymbol{M}(\boldsymbol{x})x^{m}\ mod\ g(x)] \tag{5}$$

使用彼得森-戈伦斯坦-齐勒解码器(Peterson-Gorenstein-Zierler decoder，PGZ))解码，接收到的码字 $\boldsymbol{r}(\boldsymbol{x})=\{r_0,r_1,\cdots r_{n-1}\}$，则码字多项式为：

$$\boldsymbol{r}(\boldsymbol{x})=\sum_{i=0}^{k-1}r_i x^{i} \tag{6}$$

$\boldsymbol{X}_i=\alpha^{I_i}$ 为错码位置上的变量 x 项，$\boldsymbol{Y}_i=er_i$ 为错码位置上系数项。根据误差多项式(7)定义典型值 $\boldsymbol{s}_i=r(\alpha^i)$ 后使用定位多项式(8)查找错误位置 $\boldsymbol{I}_0,\boldsymbol{I}_1,\cdots,\boldsymbol{I}_{v-1},0\leqslant\boldsymbol{I}_i\leqslant n$，根据错误位置使用式(9)计算误差值后复原始信息。

$$e(x)=r(x)-s(x)=\sum_{i=0}^{n-1}e_i x^{i} \tag{7}$$

$$\Lambda(x)=\prod_{i=0}^{v-1}1-X_i x \tag{8}$$

$$\boldsymbol{r}'_{I_i}=\boldsymbol{r}_{I_i}-\boldsymbol{Y}_i \tag{9}$$

同样，n 个接收器的情况下的原始信息表示为：

$$\boldsymbol{r}'_{I_i j}=\boldsymbol{r}_{I_i j}-\boldsymbol{Y}_{ij} \tag{10}$$

神经网络的总损失函数 L 定义为原始网络的损失和秘密信息嵌入的损失两部分的加权和。L_M 是原始网络的损失，L_r 是秘密信息嵌入的损失。β 用于调整两部分权重。使用式(14)计算提取误差。

$$L=\beta\cdot L_M+L_r \tag{11}$$

$$L_M=\frac{1}{n}\sum_{i=0}^{n-1}\|\boldsymbol{P}_{Mr}-\boldsymbol{M}_r\|^2 \tag{12}$$

$p_{m_r}(i)\in[0,1]$，$m_r(i)$ 为二进制(取值为“0”或“1”)。所以 $m_r(i)=1$ 时，$p_{m_r}(i)>0.5$ 即可保证正确提取。提取的秘密信息使用 RS 码纠错。以胶囊网络为例，该方案具有良好的嵌入容量和鲁棒性。

2) 胶囊网络架构

胶囊网络(capsule networks，CapsNets)的雏形于 2011 年提出[21]，并于 2017 年通过协议路由机制得到丰富[22]。每个胶囊是一组神经元，每个神经元表示一个实体的某些属性，如姿势、形变和速度等。较低层的胶囊使用变换矩阵对较高级的胶囊进行姿态预测。动态路由通过迭代更新保持实体的变化。本文提出的胶囊网络的基本架构如图 2 所示。网络配备了 CapsNets 和 3 个接收者卷积层。CapsNets 包括输入层、卷积层(conv1～2)、主胶囊层(PrimaryCaps)和数字胶囊层(DigitCaps)，在网络模型训练过程中嵌入秘密信息。3 个接收者卷积层(conv4～6)使用特定密钥提取秘密信息；conv1 和 2 中均有 256 个卷积核，大小为 9×9，步长为 1，Relu 激活，用于提取低级特征。主胶囊层是一个包含 32 个胶囊的卷积层胶囊，每个胶囊包含 8 个大小为 9×9 的卷积核，步长为 2，Relu 激活。数字胶囊层包含 10 个大小为 16D 的胶囊。主胶囊层和数字胶囊层之间执行动态路由，增强积极特征，忽略消极特征。接收者使用 3 个卷积层 conv4～6 分别包括 64，32，16 个卷积核，大小为 3×3，步长为 1，Relu 激活。使用嵌入密钥产生解码网络参数，这些参数在训练过程中保持不变。接收者接收到的数据使用 RS 码进行纠错。系统编码将原始消息视为多项式的系数序列，在编码后的码字后附加纠错码数据。使用 PGZ 解码器进行解码。查找到错误位置后，使用高斯-乔丹消去算法计算误差值后恢复原始消息。

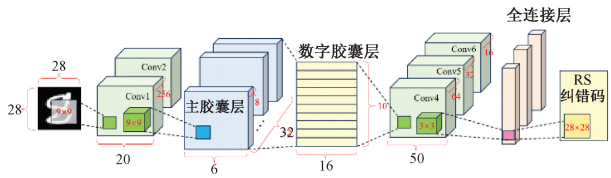


图 2　胶囊网络架构

1.2　信息嵌入与提取

本文提出的数据隐藏框架中，嵌入过程旨在通过基于矩阵乘法的矩阵编码将秘密信息嵌入到载体中。神经网络的输出用 $\boldsymbol{u}_i,i\in\{1,2,\cdots n\}$ 表示，通过将 $\boldsymbol{u}_i$ 乘以大小为 8×16 的权重矩阵 $\boldsymbol{W}_{ij}$ 来计算预测向量 $\overline{\boldsymbol{u}}_{j|i}$。将预测向量 $\overline{\boldsymbol{u}}_{j|i}$ 作为覆盖序列，通过矩阵乘法将秘密信息隐藏到覆盖序列中。

$$\begin{cases}\overline{\boldsymbol{u}}_{j|i}H_1=\boldsymbol{P}_{M1}\\ \overline{\boldsymbol{u}}_{j|i}H_n=\boldsymbol{P}_{Mn}\end{cases} \tag{13}$$

其中，$\{\boldsymbol{H}_1,\boldsymbol{H}_2,\cdots,\boldsymbol{H}_n\}$ 是嵌入矩阵，可由 $\{\boldsymbol{K}_1,\boldsymbol{K}_2,\cdots,\boldsymbol{K}_n\}$ 计算。$\{\boldsymbol{M}_1,\boldsymbol{M}_2,\cdots,\boldsymbol{M}_n\}$ 可通过式(12)中的矩阵运算和舍入运算提取。矩阵乘法可以通过全连接层实现，嵌入数据视作全连接层参数。

接收者使用卷积接收数据后输入全连接层解码。与全连接层相反，卷积层的输出是从多个信道和卷积核操作获得的复合结果，旨在保持输入和输出的多维形状一致，保留数据的重要空间信息，使网络更好地理解具有形状的数据。因此，数据解码网络被设计为 n 个卷积层和覆盖元素连接，经过卷积层后进入全连接层解码。每个全连接层的激活函数设置为 sigmoid，以将 $\{\boldsymbol{P}_{M1},\boldsymbol{P}_{M2},\cdots,\boldsymbol{P}_{Mn}\}$ 的数值限制在[0,1]中。卷积层和全连接层的参数通过使用相应的嵌入密钥生成，而不是训练。因此，解码网络参数在训练过程中保持不变。嵌入的密钥可以通过带有安全密钥的信道传输[23]。没有嵌入密钥就不能提取秘密数据，因为网络的预测向量参数未知。并且，数据隐藏密钥的信息量是远小于秘密信息的信息量的。生成卷积层的参数后，使用损失函数训练网络以保证网络模型的参数和秘密信息的正确提取。

2　实验结果

本文通过对 CapsNets 进行了大量实验验证了提出方案的有效性。方案从嵌入容量和鲁棒性两方面进评估。

2.1　实验环境

1)环境设置。所有实验均由 Tensorflow 实现，并在

3.9 环境下 3th Gen Intel(R) Core(TM) i7-1360P CPU 的 Windows 11 系统上进行训练。

2)数据集。实验采用(修改后的国家标准与技术研究所数据库)对神经网络模型执行分类任务。MNIST 包括 60 000 个大小为 28×28 的手写数字图像用于训练,10 000 张同样大小的手写数字图像用于测试。

3)对比方法。将本文提出的抗鲁棒性模型隐写方法与多接收器神经网络算法进行了多接收器在均匀噪声和高斯噪声有损信道传输的抗鲁棒性能比较。

4)评价指标。采用检测精度和提取误差进行质量评估。检测精度是秘密信息正确嵌入的客观指标。提取误差用于评估是否正确提取秘密信息,提取误差 e 的值在 0～1 之间,$e=0$ 时表示可以正确提取。提取误差的定义为:

$$e=\frac{1}{nt}\sum_{r=1}^{n}\sum_{i=1}^{t}\left|round(p_{m_r}(i))-m_r(i)\right| \tag{14}$$

2.2 参数确定

网络的总损失 L 被定义为两个部分的加权和,如式(11)所示。L_r 是原始网络损失,用于保证数据集上满意的检测精度,L_M 用于促进附加数据的嵌入,β 用于调整两部分的权重,平衡检测精度与数据隐藏的性能。较大的 β 值有利于数据提取,不利于检测精度。

为了确定 β 的取值,在一组容量为 700,迭代次数为 2,批量为 50 的 MNIST 数据集上,对一个接收者的情况进行实验。图 3 展示了包含附加数据的网络检测精度和数据提取误差。如图 3(b)所示,$\beta>1.0$ 时,检测精度明显降低,在 $\beta=1.0$ 时,提取误差如图 3(a)所示,提取误差表现良好。综合考虑取定 $\beta=1.0$,以保证满意的检测精度和提取误差。

2.3 嵌入容量

嵌入容量是数据隐藏最重要的指标之一。为验证本文提出方案能提高网络模型嵌入容量,本文在 CapsNets 中嵌入秘密信息,在 MNIST 数据集上分别针对一个接收器和多个接收器的情况测试嵌入容量。实验损失项采用交叉熵作为损失函数。Adam 优化器针对 MNIST 数据集进行优化。

对于一个接收器的情况($n=1$),秘密信息长度被设置为{500,1 000,…,8 500}。MNIST 数据集上的提取误差和检测精度在图 4 中展示。如图 4(a)所示,当容量 $t<7\ 500$ 时,可以正确提取秘密信息。因此,提出方案的单接收器嵌入容量是 7 500 bits。文献[20]中多接收器神经网络单接收器嵌入容量为 4 500 bits,提出方法嵌入容量增大了 66.6%。如图 4(b)所示,检测精度平均提高 0.17%。随着嵌入容量的增大,隐写神经网络的检测精度略有下降。隐写神经网络的检测精度嵌入数据增大过程中,由于网络参数的冗余,神经网络检测精度略有下降。这是由于隐写术和原始任务之间存在权衡。只要 β 值足够小,使网络一致收敛,可以保证嵌入的秘密数据的网络精度是令人满意的。

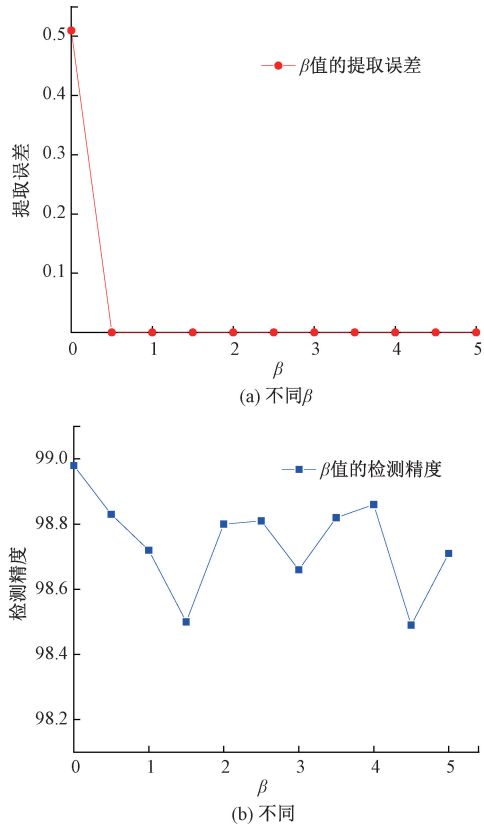


图 3 不同 $\boldsymbol{\beta}$ 值的提取误差和检测精度

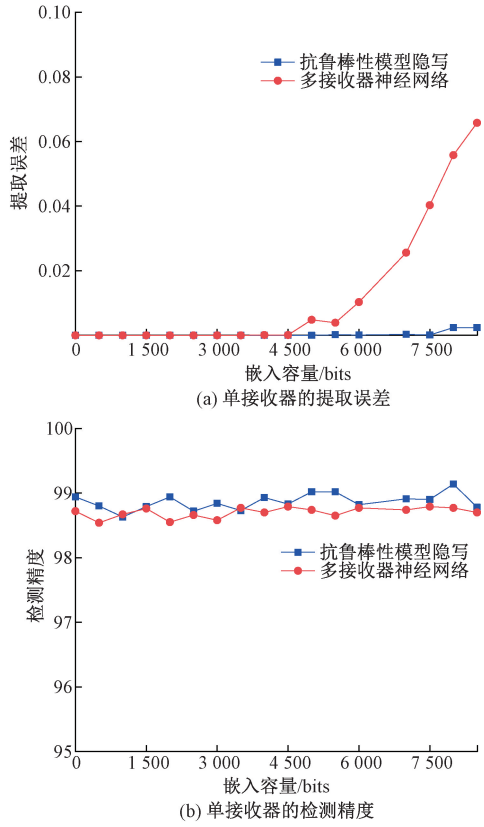
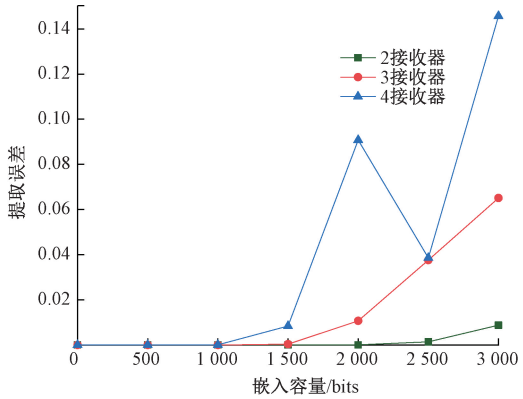


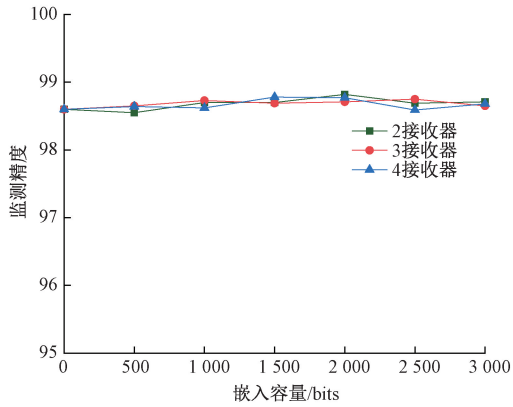
图 4 单接收器的提取误差和检测精度

可以通过训练找到既满足网络精度又满足提取误差的 β 值。图 4 中的结果表明，实验选取的 $\beta = 1.0$ 合理，随着嵌入容量的增加，检测精度是稳定的。

对于多个接收器的情况，每个接收者使用正确密钥可以提取对应部分的秘密信息。秘密信息长度被设置为{500，1 000，…，3 000}和{500，1 000，…，4 000}。类似的，MNIST 数据集上多接收器神经网络的检测精度和提取误差率以及提出方案的检测精度和提取误差率在分别在图 5、6 中示出。图 5、6 中可以看出嵌入容量的下降，这是因为多个接收机总嵌入容量被分为多个部分，每个接收机的容量小于单个接收机容量。如图 5(a)所示，在能够正确提取秘密信息的情况下，文献[20]中多接收器神经网络 2、3、4 个接收器的嵌入容量分别为 200、1 500、1 000。如图 6(a)所示，提出方案对于 2、3、4 个接收器的嵌入容量分别为 4 000、2 500、2 000。同样提出方法嵌入容量增大了 66.6%。多个接收者将总容量分成多个部分，并且所提出的方案对于神经网络中的总嵌入容量保持不变。



(a) 多接收器神经网络多接收器的提取误差



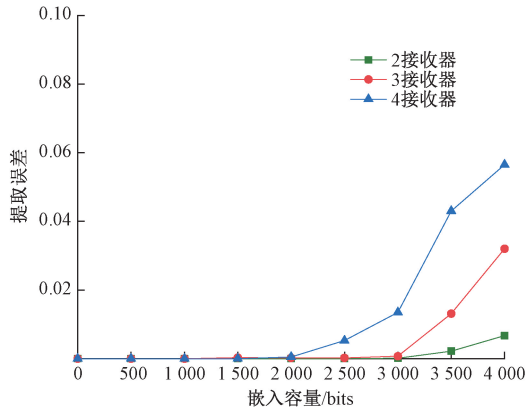
(b) 多接收器神经网络多接收器的检测精度

图 5　多接收器神经网络多接收器的提取误差和检测精度

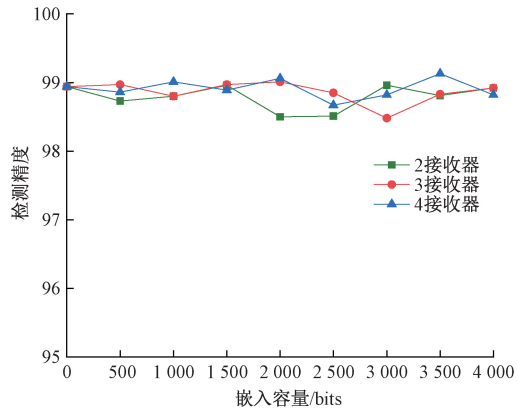
本方案与文献[20]中的深度网络模型比较，嵌入容量有明显提升。本文提出方案的嵌入容量高于文献[20]提出的多接收器神经网络。

2.4　鲁棒性

实际的通信信道常是有损耗的，接收器获得的隐写网络可能包含噪声。数据解码网络输出矢量 $\boldsymbol{P}_M$ 会包含噪声。为了测试该方案的鲁棒性，将高斯分布和均匀分布的噪声分别添加到 $\boldsymbol{P}_M$ 中，以模拟有损信道噪声。其中，高斯噪声分布服从 $N(0,\phi_1^2)$，均匀分布服从 $U(0,\phi_2)$，取 $\phi_1 = 0.14$，$\phi_2 = 0.24$。在 python 中通过 np.random.normal() 和 np.random.rand() 函数生成。采用两组对比实验观察网络模型鲁棒性。MNIST 数据集上的提取误差率在图 7、8 中示出。如图 7(a)、(b)所示，与文献[20]中提出的方法相比，单接收器使用 RS 码纠错前后提取误差，在高斯噪声情况下，提取误差平均减小 39.04%，检测精度提高 0.26%。

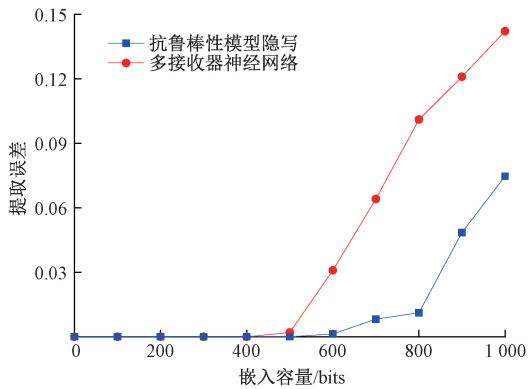


(a) 提出方案多接收器的提取误差

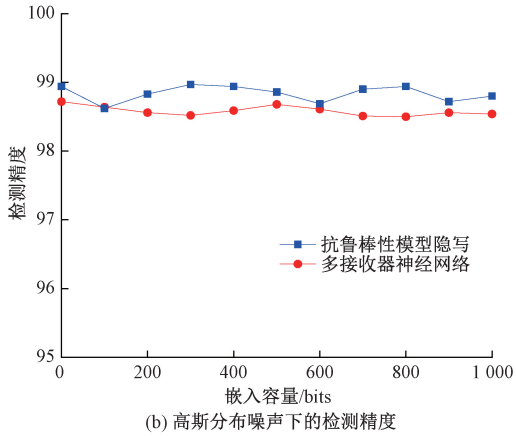


(b) 提出方案多接收器的检测精度

图 6　提出方案多接收器的提取误差和检测精度



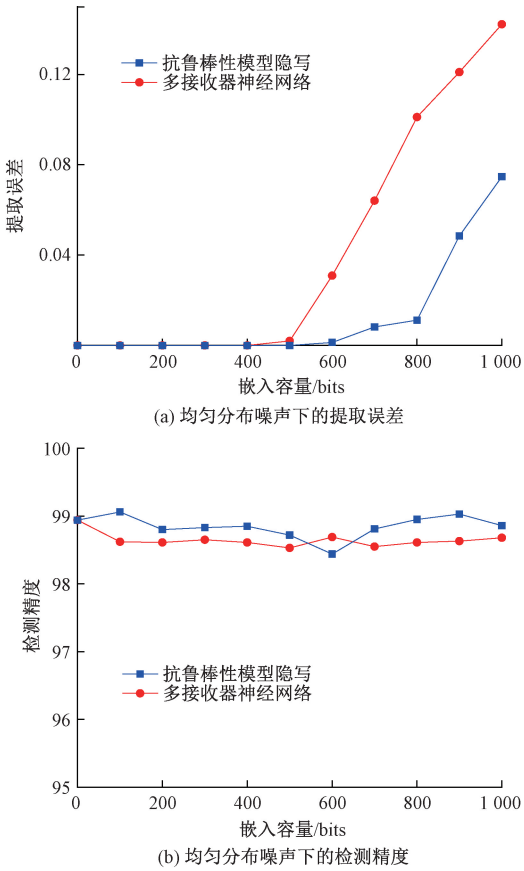
(a) 高斯分布噪声下的提取误差



(b) 高斯分布噪声下的检测精度

图 7 高斯分布噪声下的提取误差和检测精度

如图 8(a)、(b)所示,在均匀噪声下,提取误差平均减小 79.9%,检测精度提高 0.20%。推广至多接收器,相同的嵌入容量下,使用 RS 码纠错前后提取误差,在高斯噪声情况下,提取误差减小。在均匀噪声下提取误差减小。图 8 中可以看出随着嵌入容量的增大,鲁棒性降低,这是由于更大的嵌入容量减少了数据的冗余。



(a) 均匀分布噪声下的提取误差

(b) 均匀分布噪声下的检测精度

图 8 均匀分布噪声下的提取误差和检测精度

本文提出的方案在 CapsNets 中具有更高的嵌入容量的同时,提高了提取秘密信息的鲁棒性。实验结果表明提出方案具有良好的鲁棒性。

3 结 论

针对嵌入容量和提取鲁棒性低的问题,本文提出了一种以神经网络模型为载体的鲁棒隐写方法。首先在训练过程中将秘密信息嵌入到神经网络中,而不是在训练结束后修改网络参数,训练过程中网络参数保持不变。之后接收者使用正确密钥提取相应部分的秘密信息。解码网络参数由接收器的拥有嵌入密钥生成,接收到的数据使用 RS 码纠错。实验结果表明提出的抗鲁棒神经网络模型能提升信息隐藏的嵌入容量和鲁棒性。使用 CapsNets 网络模型时,本文提出的方法优于现有的网络模型隐写方案。未来进一步的研究中,可将提出的方案应用于更多的神经网络以实现信息隐藏的更大的嵌入容量和更好的鲁棒性。

参考文献

[1] 孟若涵. 载体增强式信息隐藏方法研究[D]. 南京:南京信息工程大学, 2023.

[2] WANG Z, ZHANG X, YIN Z. Joint cover-selection and payload-allocation by steganographic distortion optimization[J]. IEEE Signal Processing Letters, 2018, 25(10): 1530-1534.

[3] LI S, JIA Y, KUO C C J. Steganalysis of QIM steganography in low-bit-rate speech signals[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 25(5): 1011-1022.

[4] RABIE T, BAZIY A M. The pixogram: Addressing high payload demands for video steganography[J]. IEEE Access, 2019, 7(2): 21948-21962.

[5] FI L T, JU D J, FRI D J. Minimizing additive distortion in steganography using syndrome-trellis codes[J]. IEEE Transactions on Information Forensics and Security, 2011, 6(3): 920-935.

[6] HOLUB V, FRIDRICH J, DENEMARK T. Universal distortion function for steganography in an arbitrary domain[C]. EURASIP Journal on Information Security, 2014, 2014(1): 1-13.

[7] VAHID S, COGRANNE R, FRIDRICH J. Content-adaptive steganography by minimizing statistical detectability[J]. IEEE Transactions on Information Forensics and Security, 2015, 11(2): 221-234.

[8] WU P, YANG Y, LI X Q. Mega image steganography capacity with deep convolutional network[J]. Future Internet, 2018, 10(6): 54-68.

[9] HU D H, WANG L, LIANG W J, et al. A novel image steganography method via deep convolutional generative adversarial networks[J]. IEEE Access, 2018, 6(1): 38303-38314.

[10] ZHANG C, BENZ P, KARJAUV A, et al. Udh:

Universal deep hiding for steganography, watermarking, and light field messaging[C]. Neural Information Processing Systems, 2020, 33(1): 10223-10234.

[11] ELHAKI O, SHOJAEI K. Neural network feedback linearization target tracking control of underactuated autonomous underwater vehicles with a guaranteed performance[J]. Ocean Engineering, 2018, DOI: 10.1016/j.oceaneng.2018.08.007,167: 239-256.

[12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. IEEE Conf. Comput. Vision Pattern Recognit, 2016(3): 770-778.

[13] 郑乐佳. 基于卷积神经网络的分数像素运动补偿[J]. 电子测量技术, 2023, 46(13): 124-131.

[14] WANG Z, ZHANG X, YIN Z. Joint cover-selection and payload-allocation by steganographic distortion optimization[J]. IEEE Signal Process, 2018,25(10): 1530-1534.

[15] CHEN H, SONG L, QIAN Z, et al. Hiding images in deep probabilistic models[J]. Neural Information Processing Systems, 2022. 6(7): 2210-2219.

[16] ADI Y, BAUM C, CISEE M, et al. Turning your weakness into a strength: Watermarking deep neural networks by backdooring[J]. Security Symp, 2018, 18(27): 1615-1631.

[17] YANG Z, WANG Z, ZHANG X, et al. Multi-source data hiding in neural networks[J]. IEEE International Workshop on Multimedia Signal Processing, 2022, 10(24): 1-6.

[18] ZHU J, KAPLAN R, JOHNSON J, et al. Hidden: Hiding data with deep networks[J]. Proceedings of the European Conference on Computer Vision (ECCV), 2018, 3(25): 657-672.

[19] GOODFELLOW I J, POUGET-ABADIE J, MRZA M, et al. Generative Adversarial Networks[C]. Neural Information Processing Systems. Montreal,2014.

[20] WANG Z, FENG G, WU H, ZHANG X. Data hiding in neural networks for multiple receivers[J]. IEEE Computational Intelligence Magazine, 2021,16(4): 70-84.

[21] HINTON G E, KRIZHEVSKY A, WANG S D. Transforming auto-encoders [C]. International Conference on Artificial Neural Networks, 2011.

[22] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[J]. Neural Information Processing Systems 2017, 30(26): 3856-3866.

[23] SIMMONS G J. The prisoners′ problem and the subliminal channel[J]. Proceedings of CRYPTO ′83, 1984, 83(19): 51-67.

作者简介

杨彤彤(通信作者),本科,主要研究方向为隐写。

E-mail:ytt0322@outlook.com

杨紫云,硕士研究生,主要研究方向为隐写。

E-mail: daxiabbq@163.com

王子驰,博士,副研究员,主要研究方向为隐写、隐写分析、人工智能安全等。

E-mail: wangzichi@shu.edu.cn