文章编号:1674-2974(2022)04-0018-08 DOI:10.16339/j.cnki.hdxbzkb.2022278

一种基于深度学习的移动端隐写方法

廖鑫[1],黎懿熠[1,2],欧阳军林[2],周江盟[3],戴湘桃[4],秦拯[1†]
(1. 湖南大学 信息科学与工程学院,湖南 长沙 410082;
2. 湖南科技大学 计算机科学与工程学院,湖南 湘潭 411201;
3. 中南大学 物理与电子学院,湖南 长沙 410083;
4. 长城信息股份有限公司,湖南 长沙 410199)

摘 要:隐写是隐蔽通信的主流方法之一,而移动端则是当下最常用的通信设备,二者的结合研究具有较高的实际意义.近年来,基于深度学习的隐写方法得到快速发展,然而在性能提升的同时,各类网络结构向着更复杂、庞大的方向演变,逐渐脱离以隐蔽通信为核心的实际应用场景,实用性较低.针对这一现象,本文提出一种适用于移动端的轻量级图像隐写方法.对网络整体进行轻量化设计,结合深度可分离卷积降低模型计算量,在精度和速度之间取得较好的折中平衡.以生成对抗网络的思想,将编码器、解码器和判别器构成的整体模型纳入对抗训练中,使子网络在迭代对弈中实现螺旋式上升发展.为应对真实环境下的各类挑战,模型被落地部署于移动设备上进行真机实验.在移动端,精简后的模型性能会出现小幅下降.对此,在方法中引入BCH纠错码以确保正确提取信息.实验结果表明,该移动端隐写方法生成图像质量好,且具有较高的响应速度,能满足现代社会中人们对便捷性的高要求.值得注意的是,该方法的所有计算工作均可在移动端独立完成,不需要通过网络请求服务器,能避免网络窃听攻击.

关键词:隐写;深度学习;生成对抗网络;移动端;轻量级
中图分类号:TP309 **文献标志码**:A

A Mobile Steganography Method Based on Deep Learning

LIAO Xin[1], LI Yiyi[1,2], OUYANG Junlin[2], ZHOU Jiangmeng[3], DAI Xiangtao[4], QIN Zheng[1†]
(1. College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China;
2. School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China;
3. School of Physics and Electronics, Central South University, Changsha 410083, China;
4. Great Wall Information Co., Ltd, Changsha 410199, China)

Abstract: Steganography is one of the main methods for covert communication, while mobile phones are the most commonly used communication devices. The combination of the two has high practical significance. In recent

* **收稿日期**:2021-11-29
基金项目:国家自然科学基金资助项目(61972142, 61772191, U20A20174), National Natural Science Foundation of China (61972142, 61772191, U20A20174);国家社会科学基金资助项目(21BXW077), National Social Science Foundation of China (21BXW077);湖南省重点研发计划项目(2019WK2072), Key Research and Development Program of Hunan Province (2019WK2072);湖南省自然科学基金资助项目(2020JJ4212, 2021JJ30277), Natural Science Foundation of Hunan Province (2020JJ4212, 2021JJ30277)
作者简介:廖鑫(1985—),男,湖南长沙人,博士,副教授,博士生导师
† 通信联系人,E-mail: zqin@hnu.edu.cn

years, steganography has developed rapidly with deep learning technologies. To improve the performance, networks evolve towards a more complex and large style, which gradually deviates from the real world scenarios with covert communication as the core, resulting in low practicability. For convenience and efficiency, a lightweight image steganography method is proposed for mobile phone. The network structure is designed in a light style, with depthwise separable convolutions utilized to reduce useless parameters and keeping a balance between accuracy and speed. Based on generative adversarial networks, the proposed method consists of a generator, a decoder, and a discriminator, which are trained together defiantly and finally advance in a spiral upward trend. To deal with various challenges in the real world, the model is deployed on mobile phones for tests. The networks used on smartphones are pruned, which indicates performance degradation. To ameliorate this problem and enhance decoding accuracy, BCH correcting codes are used in the method. The results show that the method can generate high-quality images with high speed, which meets the convenience requirements in today's world. Besides, it's worth noting that the method works without online requests. All the embedding and extracting tasks can be done by phone itself, which means this scheme is immune to eavesdropping attacks.

Key words: steganography; deep learning; GAN; mobile phone; lightweight

信息时代下,数据背后的价值被挖掘显现,人们逐渐意识到信息安全的重要性,并对其提出了更高的要求.在众多信息保护方法中,隐写技术[1]不仅能保障信息本身的安全性,也使得秘密信息的传递过程不易被感知.通过隐写技术,秘密信息被嵌入图像、音频和视频等多媒体载体[2-3].这些多媒体内容在嵌密前后几乎无异,人类无法感知其间细微的变化.如此,秘密信息即可随着图像音频等介质一并传播,实现隐蔽且安全的传输.

隐写兼具安全性和隐蔽性的双重特性,逐渐成为信息安全领域中的一大研究热点[4-8].目前已经出现了许多围绕隐写开展的研究[9],然而其中鲜有针对移动端设计的隐写方法.近年来,深度学习技术被广泛应用于隐写领域,为进一步挖掘数据潜力,模型被设计得愈发复杂多变,其训练和使用需占用更多计算资源.这些大型网络难以被应用于移动设备.另外,应用市场中的隐写应用大多是基于早期方法的简单实践[10],譬如基于 LSB 算法的 PocketStego 和 Steganography_M.这些应用的安全性较低,难以抵抗最常见的攻击,且极易被侦破,亟须发展针对移动平台的高可用性隐写技术.

隐写技术常用于隐蔽通信,而移动设备则是当今社会中最常用的通信设备,二者的结合研究显得理所当然.早期移动设备算力的不足阻碍了相关研究的开展,直到现在硬件技术的大幅度提升才为这一思路提供了可行平台.本文提出了一种基于生成对抗网络[11]的移动端隐写方法,通过对抗训练的方式逐步提高模型的隐写效果,能将秘密信息嵌入自然图像中,输出包含秘密信息的载密图像.

本文的主要贡献如下.

1)基于对抗的思想,设计由编码器、解码器和判别器构成的整体模型框架,三者在对抗训练中相互博弈,通过合理的损失函数进行约束,最终呈螺旋上升式进步,编码器生成图像质量提升,同时解码器能更准确地提取并还原秘密信息.

2)以有效和精简为原则进行网络模型设计,在性能和轻量之间取得较好的折中,减少编码器和解码器网络层数,同时使用深度可分离卷积进一步减少模型的计算量.与现有的其他深度学习隐写方法相比,所提出的方法资源占用量更低,且能维持较好的性能.

3)以实用性为出发点,对所提出的方法进行落地实现,并在实际应用场景中进行真机测试.对于跨平台算子改变造成的模型性能下降问题,结合 BCH 纠错码提高解码正确率,保障了信息的可靠传输,验证了基于深度学习的隐写算法在移动端平台的可行性.

1　相关工作

1.1　生成对抗网络

Goodfellow 提出的生成对抗网络(Generative Adversarial Networks, GAN)[11],是深度学习中的一种重要算法.一个典型的 GAN 模型通常包含生成器和判

别器.生成器的目的在于使得生成图像的分布与自然图像尽可能类似,让人无法用肉眼辨别.而判别器的任务则是辨别输入图像是否为生成图像,其目标和利益恰好与生成器相反.这二者交替进行训练,其中一方或会率先取得进步,但不久后另一方便会追赶上来.二者的进步迭代交叉,呈螺旋上升的趋势.

1.2 传统隐写与深度学习隐写

传统隐写技术可以分为自适应和非自适应两类.一些非自适应的隐写方法利用信道编码中的技术实现矩阵嵌入,譬如一些基于汉明码和基于方向编码的矩阵嵌入方法.此外还存在修改区域可选的矩阵嵌入方法,如湿纸码[12].与非自适应隐写方法不同,自适应隐写方法会依据图像内容,有针对性地选择纹理丰富度更高的区域进行嵌密.自适应隐写方法常依靠最小失真框架实现,譬如经典的HUGO[13]和UNIWARD[14]等.嵌入失真代价函数和STC[15]构成了该框架的主要组成部分.

近年来,深度学习兴起并进入了快速发展的阶段[16-18],其相关技术被引入隐写领域,与传统隐写方法碰撞出一些新的研究方法.Hayes等人结合深度学习,提出包含Alice、Bob和Eve三个子网络的隐写模型[19],输出的载密图像与原始图像具有较高相似性.Zhu等人侧重考虑鲁棒性,在所提出的HiDDeN[20]中增加噪声层,用以模拟JPEG压缩和各类噪声,从而提升鲁棒性,但其存在嵌密量较低的缺点.Bernard等人提出的方法中[21],隐写方可以使用对抗样本和动态STC等工具用以嵌密,而隐写分析方则拥有许多分析方法,双方模拟对抗游戏,使得模型收敛并获得一种高效的隐写算法,充分挖掘了经典隐写方法的安全性.

1.3 移动端隐写应用

目前的研究中,针对移动设备的较少,且现有的移动端应用大多是基于传统方法的实践[10],譬如LSB算法及其变型和F5算法等.这些移动端隐写方法不足以抵挡现在的隐写分析技术,存在安全性较低的问题.在自适应隐写上,Su等人[22]提出了基于J-UNIWARD的移动端隐写,其安全性有所提高.近年来出现的基于深度学习的隐写方法大多存在体量较大的问题.尽管随着硬件设备的发展,移动设备的算力提升且内存增加,但这些大体量的隐写方法仍会占用设备的大量资源,无法满足移动端及时响应的需求.

2 基于GAN的移动端隐写方法

2.1 模型结构

隐写分析是与隐写对立的技术,可用于分析图像中是否包含秘密信息[23].隐写与隐写分析之间的对抗和GAN中生成器与判别器之间的对抗十分相似,因此GAN能被自然地运用到隐写任务中.所提出的移动端图像隐写方法综合考虑了发送者和接收者的应用需求,由生成器、解码器和判别器三个部分组成,如图1所示.向生成器输入载体图像I_C和待嵌入信息,经处理后即可输出载密图像I_S.解码器的目的则是从载密图像I_S中提取并还原秘密信息.判别器则发挥着隐写分析的作用.

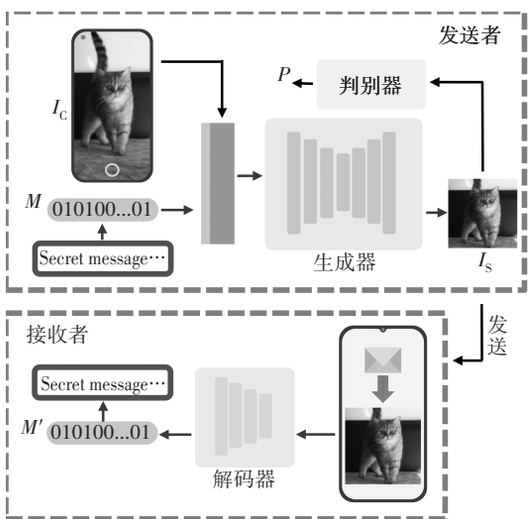

图1 基于GAN的移动端隐写方法整体结构

Fig. 1 Structure of mobile steganography based on GAN

拍摄的原始图像和秘密信息需经历一番转换才能输入生成器.首先将秘密信息按ASCII码转换为01串,然后将其变形为三维矩阵并扩展为与载体图像相同长宽的形状,这样便能在通道维度将二者拼接为一个整体.此处的扩展操作通过反卷积实现,而非全连接,这样可以很大程度上减少计算量.生成器对该整体提取特征,从而获得载密图像.生成器的网络结构启发自U-Net[24],先利用卷积进行下采样,逐步缩小特征图大小,并获得多尺度特征;而后进行反卷积上采样,同时级联之前获得的特征,输出特征图逐渐恢复为原始大小,并在最后作为残差图像输出.将残差图像与载体图像相加,即可得到载密图像.该模型较好地结合了浅层、深层信息,有利于更好地感知图像信息,从而生成图像质量更好的载密图像.考

虑到应用场景是移动端,在部署模型时,可将模型中的标准卷积更换为深度可分离卷积,从而减小网络计算量.不同嵌密量的生成器网络有着微小的区别,在图2中可以看到20 000 bit嵌密量下的生成器结构.

解码器网络是类似漏斗形的结构,它由多层卷积组成,每层输出特征图大小逐渐减小,提取关键特征以还原秘密信息,并期望最后解码的信息与原始秘密信息M尽可能一致,如图2所示.在训练阶段本文使用随机01串模拟秘密信息,在部署运用时则结合BCH纠错码以提高解码正确率.

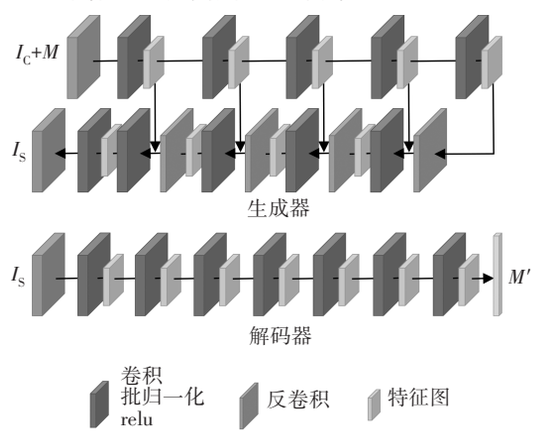

图2　生成器和解码器的网络结构

Fig. 2　Generator and decoder networks

判别器作为一个二分类器,发挥了类似隐写分析器的作用,目的是督促生成器输出图像质量更好的载密图像,提高安全性.在本工作中,使用Ye-Net[25]作为判别器.Ye-Net是2017年由Ye等人提出的基于深度学习的隐写分析器,它是现在最先进的隐写分析方法之一.

2.2　损失函数

生成器与解码器作为一个小整体,与判别器交替进行训练,二者相互促进,直到达到纳什均衡点.判别器的损失函数设计如下:

$$L_{\text{Discriminator}} = \mathbb{E}_{X \sim P_C}\left(\text{Ye - Net}\left(I_S\right)\right) \tag{1}$$

式中:I_S表示载密图像,Ye-Net(I_S)表示载密图像通过Ye-Net后得到的评判结果.解码器使用交叉熵损失函数进行约束,损失函数表示如式(2).

$$L_{\text{Decoder}} = \mathbb{E}_{X \sim P_C}\text{CrossEntropy}\left(M_i, M_i'\right) \tag{2}$$

式中:M_i表示秘密信息,M_i'表示提取的秘密信息,CrossEntropy是交叉熵损失计算.生成器的损失函数包括图像损失和安全性评估,如式(3)~式(5).

$$L_{\text{Generator}} = \alpha L_{\text{I}} + \beta L_{\text{S}} \tag{3}$$

$$L_{\text{I}} = 1 - 0.5 \times \left(\tau\text{PSNR}\left(I_C, I_S\right) + \text{SSIM}\left(I_C, I_S\right)\right) \tag{4}$$

$$L_{\text{S}} = \mathbb{E}_{X \sim P_C}\text{CrossEntropy}\left(\text{Ye - Net}\left(I_S\right), \text{Ye - Net}\left(I_C\right)\right) \tag{5}$$

式中:I_C表示载体图像,PSNR(I_C, I_S)和SSIM(I_C, I_S)分别表示载体图像与载密图像之间的峰值信噪比和结构相似性,τ、α以及β是计算参数,数值分别为0.01、1和0.5.

通过损失函数约束,生成器网络逐步趋向于输出质量更好的载密图像,它与载体图像在肉眼观察中难以进行区分,且不易被隐写分析器检测出来.同时,解码器的解码能力亦趋向于提升.

2.3　轻量化处理

本工作的应用场景为移动端,这要求网络模型在满足精度的同时,尽可能轻量化.本文尝试利用所提出的隐写方法构建安卓应用,所得到的APK安装包大小约20.9 MB.以同样的方法尝试对StegaStamp[26]和Hayes等人的方法[19]进行构建,获得了大小分别为400 MB和848 MB的APK安装包.由此可以直观地感受到所提出的模型具有较高的轻巧性.

此外,本文还针对资源不足的设备,提出了进一步的轻量化改进方法.实验中将标准卷积替换为深度可分离卷积,以期减少模型计算量.深度可分离卷积采用先分后合的结构,首先按单个通道分别进行卷积,这样能大大减少所需的参数量;而后将这些输出合并再执行1×1卷积.后者可在通道维度上进行学习,弥补了单通道卷积的不足.

优化后的模型可以从文件大小上直观感受到模型体量明显下降.包含有生成器与解码器的模型文件由原先的4 087 kB缩小到了2 122 kB,APK文件则从20.9 MB缩小到14.7 MB.进一步地,对模型的计算复杂度进行量化评估.对于深度可分离卷积,其计算量可以表示为式(6).

$$m \times L_k \times L_k \times L_{\text{out}} \times L_{\text{out}} + m \times n \times L_{\text{out}} \times L_{\text{out}} \tag{6}$$

式中:m和n分别表示输入和输出特征图的通道数,L_k和L_{out}分别是卷积核和输出特征图的边长.轻量化后的隐写方法体量明显降低,且计算量减少到原本的19.24%.在低算力的移动设备上,可以选择部署轻量化后的隐写模型,这将有益于减少设备的计算压力,且将性能维持在较好的水平;而对于资源丰富的设备,可以选择使用原始隐写模型以获得更佳的性能体验.

3 实验

3.1 实验设置

考虑到移动端相机获取的图像大多数是JPEG格式,所以本方法中没有使用隐写任务常用的无损图像数据库,而是使用mirflickr25k数据集.该数据集由25 000张JPEG格式的图像组成,囊括多种分类标签,比如clouds、male和food等.本文的实验中将数据集分为两部分,其中20 000张图作为训练集,5 000张图作为测试集.模型训练在NVDIA GeForce RTX 2080 Ti GPU上完成,并使用了Adam优化器帮助收敛.

3.2 隐写图像质量和解码准确率

将所提出的隐写方法与两个基准方法StegaStamp[26]和ReDMark[27]进行对比,其结果如表1所示.本文方法在图像质量指标SSIM和PSNR上均优于其他方法,表现出较好的性能.解码准确率也更高,达到了99.37%.在体量方面,模型文件大小与模型计算参数量相关.本文方法可对400×400像素的图像进行信息嵌入,与StegaStamp方法的输入图像大小一致.在此同一量级的对比中,本文模型更轻巧,模型文件大小仅为StegaStamp方法的五分之一.而相较于处理图像仅为32×32像素的ReDMark方法,尽管本文方法处理的图像更大,但模型中的轻量化处理使参数量大大减少,本文模型文件甚至更小一些.

表1 StegaStamp[26]、ReDMark[27]与本文方法的对比

Tab.1 Comparison with StegaStamp[26] and ReDMark[27]

隐写方法	SSIM	PSNR	准确率/%	嵌密量/bit	文件大小/kB
StegaStamp[26]	0.943 3	27.91	98.73	100	211 991
ReDMark[27]	0.977 1	40.62	98.64	6 912	4 701
本文方法	0.997 1	43.96	99.37	20 000	4 087

这里可以对上述结果作进一步的解释说明.StegaStamp方法中使用噪声层模拟各类噪声、压缩和色彩失真攻击,从而获得较高的鲁棒性.但噪声层的引入会导致一定程度的图像质量下降.相较于StegaStamp,本文的方法更注重于移动端的通信能力,选择将重点放在提升图像质量与嵌密量上.因此所提出的方案剔除了噪声层的干扰,从而使图像在视觉效果上更为清晰.文中引入PSNR和SSIM对图像质量进行评估.其中,PSNR表示信号最大值与背景噪声之间的比例大小,当噪声大幅度降低后,该指标获得显著提升.而在结构相似性方面,由于隐写任务对图像结构的改变较少,该项指标的提高不甚明显.此外,在模型体量上,本文舍弃了全连接层,结构层次也更浅,因此模型更为轻巧,甚至在提升嵌密量后亦能保持较低的模型体量.体量上的优势将有益于模型在移动端的部署,同时也有助于取得更快的响应速度.

对于深度学习方法,模型嵌密量的改变意味着模型结构的变动和性能的下降.为了维持方法原有的优良特性,同时尽可能公平地进行性能对比,本文在后续实验中将ReDMark处理的数据进行拼接处理.将多个32×32像素的处理图像拼接为384×384像素的图像,从而与处理图像为400×400像素的StegaStamp和本文方法构成同一量级.在嵌密量指标上,经过拼接处理的ReDMark方法可达到6 912 bit.根据图像质量、鲁棒性和嵌密量三指标平衡规律,信息隐藏方法的嵌密量越高,其另外两项指标将更难以提升.本文方法的嵌密量高于对比方法,在此"劣势"中进行对比,更可显现本文方法在图像质量和轻量性质中的优越性.

训练完毕并通过初步测试的模型可以借助支持库部署在移动端.首先将模型转换输出获得pb格式的模型文件,再结合从TensorFlow源码编译的libtensorflow_inference.so和libandroid_inference.jar库文件,即可在安卓端进行模型调用.在安卓工程的使用中,图像的读入读出会带来一定程度的图像质量损失,这一问题主要表现为解码准确率下降.为了缓解这一问题,本文在安卓工程中引入了BCH纠错码.输入图像类型仍是JPEG格式,而输出类型则选用PNG格式,以尽可能降低错误率.结合BCH的纠错能力,隐写应用能基本正确地还原秘密信息,且图像质量良好,见图3.

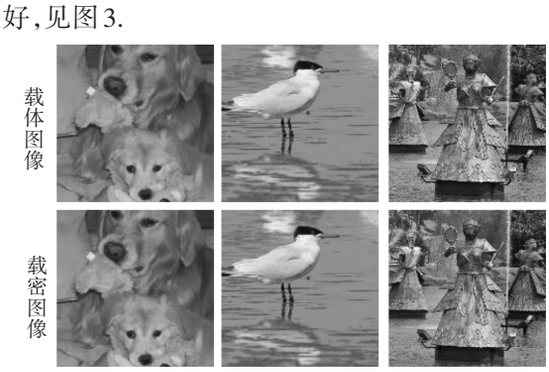

图3 移动端隐写应用中载体图像和载密图像的对比

Fig. 3 Comparison between stego and cover in proposed mobile steganography APP

3.3 模型运行时间测试

模型运行的速度与内存大小、CPU 等硬件设备相关,且不同的耗电状态和系统资源管理策略也会对计算速度带来不同程度的影响.本文通过改变安卓模拟器的 CPU 核数和内存大小,来模拟不同的 CPU 性能和内存状态,如表 2 所示.作为对比,实验中公平地将对比模型打包并转换,然后结合编译获得的支持库,以类似的安卓代码部署在移动端.在同等条件下,本文方法展现了更佳的反应速度.尤其在内存资源较为匮乏的情况下,这种优越性更为明显.在仅使用 1 核 CPU 和 2G 内存的情况下,所提出的方法比 StegaStamp 用时少了 32.4%,且远小于 ReDMark 用时. 而在 1 G 的内存下,StegaStamp 模型甚至会无法正常运行,提示内存不足然后闪退.ReDMark 方法所处理的图像较小,单次运行负担较小,但这同时带来了更多频次的处理,总体用时反而较长.

表 2　在不同情况下的模型运行时间

Tab. 2　Model runtime in different situation

		1G	2G	4G
StegaStamp[26] 用时/ms	1 核	—	5 424.2	5 347.5
	2 核	—	2 660.7	2 639.6
	4 核	—	1 545.8	1 428.6
ReDMark[27] 用时/ms	1 核	109 378.6	109 210.1	107 955.7
	2 核	52 831.3	52 157.7	49 711.8
	4 核	31 382.7	30 729.3	28 073
本文方法 用时/ms	1 核	3 666.3	3 630.7	3 891.8
	2 核	1 913.6	1 912.5	1 907.7
	4 核	1 219.8	1 114.8	1 110.2

模型的运行速度受许多因素的影响,而其中影响较大的因素是 CPU 的性能.如表 2 所示,在同等 CPU 核数下,1 G、2 G 和 4 G 内存之间的运行时间差距较小,而在不同核数下的差异较大.为进一步探讨 CPU 性能与模型运行时间之间的关系,本文将模拟器的内存固定为 4 G,同时改变模拟器的 CPU 核数进行实验,如图 4 所示.在不同 CPU 核数下,本文方法运行用时均小于 StegaStamp. 当 CPU 核数低于 4 核时,移动端算力不足,执行并发度较低,因此运行用时在很大程度上受模型计算量的影响.此状态下,计算量更低的本文方法可占据较大的优势.当 CPU 核数上升到 4 核后,设备资源相对充裕,两者的模型均能在较短时间内完成运算,因此差异较小.此状态下,两个模型的运行耗时都趋于稳定平缓.而在该稳定状态下,本文方法仍保有一定的优势.

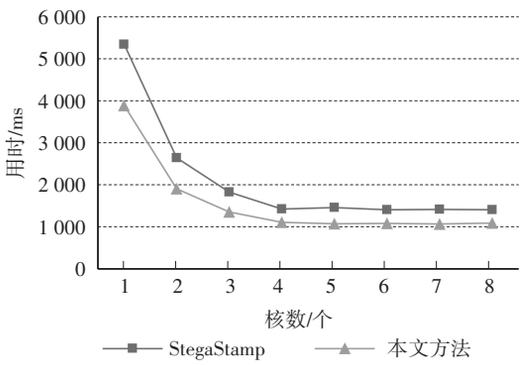

图 4　不同 CPU 核数下运行模型所需时间

Fig. 4　Model runtime with different CPU cores

3.4 运行模型的内存占用情况

除了运行时间外,本文对模型的内存占用情况进行了实验和记录.为了便于观察,本文固定在同一环境下进行多次内存占用测试.所用的安卓模拟器被设置为 1 核 CPU 且内存为 4 G.受设备状态和内存回收机制等因素的影响,测试结果常会上下浮动.在多次模型调用的过程中,内存占用往往先上升至一个峰值,然后再陡然下降并最终趋于平缓,且平缓时内存占用值比模型调用前更高一些.基于这一观察,在实验中分别统计了调用模型时的内存占用峰值和内存占用稳定值作为结果,见表 3.在移动场景下,不论是内存占用峰值还是稳定值,本文方法对内存资源的消耗均远低于 StegaStamp.其内存占用稳定值平均为 37.87 MB,完全可以满足移动设备的使用要求,甚至能在低内存的设备上正常使用.这将有助于在更轻巧的设备上部署该模型,如小型相机等.

表 3　运行模型占用的内存

Tab. 3　Memory occupation when running models

	StegaStamp[26]		本文方法	
	稳定值	峰值	稳定值	峰值
内存/MB	247.59	390.45	37.87	42.83

3.5 真机测试

在实际应用场景下,本文选用了三种常见的智能移动设备进行真机测试,包括荣耀 30、小米 10 Lite 和荣耀 9.三个移动设备的状态各不相同,其中前两者使用时长不超过一年,而荣耀 9 使用了近四年.真机实验测试的内容包括模型运行时间和解码正确率,如表 4 所示.其中,本文将图像读写带来的损失纳入考察范围,测试重载入图像的解码正确率,并记

为重载入.实验表明,所提出的隐写模型在真机实验中亦具有较好的性能,响应及时.即便是使用近四年的旧机型也能在1.7 s内输出结果,具有较高的使用普适性.

表4 真机运行模型的用时及正确率

Tab. 4 Runtime and accuracy test on real devices

	CPU	内存	用时/ms	正确率/%	重载入/%
荣耀30	华为麒麟985	6G	660	99.89	98.09
小米10 Lite	骁龙765G	8G	955	99.94	97.62
荣耀9	海思麒麟960	6G	1640	99.98	99.57

4 结 论

本文针对移动端展开研究,结合深度学习设计了一种轻量化的隐写模型.具体工作中结合生成对抗网络的思想,直接由网络输出载密图像.针对移动端的使用,对网络结构进行了改进,舍弃对全连接层的使用,并用深度可分离卷积替换标准卷积,从而降低网络体量并减少计算量.在部署时,出现解码器性能下降的问题.对此,使用BCH纠错码矫正解码错误比特,从而实现秘密信息的正确提取.实验结果表明,所提出的隐写模型在移动端具有较好的表现,能满足一般的移动端隐写需求,填补了在移动端深度学习隐写研究上的空缺.

从结果上看,所设计的隐写应用能为安全相关从业人员提供更便利的隐蔽通信方法.在未来的工作中,拟针对移动端使用时的性能下降问题开展相关研究,期望在保持较小的计算量下,尽可能提升隐写性能,尤其是生成图像的质量.这要求在轻量化中尽可能保留对隐写性能影响较大的组成部分,而删去无关紧要的分支,具体细节还有待进一步的研究.

参考文献

[1] 翟黎明,嘉炬,任魏翔,等. 深度学习在图像隐写术与隐写分析领域中的研究进展[J]. 信息安全学报,2018,3(6):2-12.
ZHAI L M, JIA J, REN W X, *et al*. Recent advances in deep learning for image steganography and steganalysis[J]. Journal of Cyber Security,2018,3(6):2-12.(In Chinese)

[2] 刘荣,李冠,贾斌. 基于改进生成对抗网络的图像自适应隐写模型[J]. 计算机工程与设计,2021,42(6):1551-1561.
LIU R, LI G, JIA B. Image adaptive steganography model based on improved generative adversarial network[J]. Computer Engineering and Design,2021,42(6):1551-1561.(In Chinese)

[3] 廖鑫, 唐志强, 曹纭. 基于生成对抗网络的空域彩色图像隐写失真函数设计方法[J]. 软件学报, 2021. DOI: 10.13328/j.cnki.jos.006290.
LIAO X, TANG Z Q, CAO Y. Steganographic distortion function learning method for spatial color image based on generative adversarial network[J]. Journal of Software, 2021. DOI: 10.13328/j.cnki.jos.006290.(In Chinese)

[4] 马媛媛,徐久成,张祎,等. 基于W2ID准则的Rich Model隐写检测特征选取方法[J]. 计算机学报,2021,44(4):724-740.
MA Y Y, XU J C, ZHANG Y, *et al*. W2ID criterion-based rich model steganalysis features selection[J]. Chinese Journal of Computers,2021,44(4):724-740.(In Chinese)

[5] 任魏翔,翟黎明,王丽娜,等. 基于卷积神经网络的JPEG图像隐写分析参照图像生成方法[J]. 计算机研究与发展,2019,56(10):2250-2261.
REN W X, ZHAI L M, WANG L N, *et al*. Reference image generation algorithm for JPEG image steganalysis based on convolutional neural network[J]. Journal of Computer Research and Development,2019,56(10):2250-2261.(In Chinese)

[6] 吴建斌,康子阳,刘逸雯,等. 基于图像分类的无载体信息隐藏方法[J]. 湖南大学学报(自然科学版),2019,46(12):25-32.
WU J B, KANG Z Y, LIU Y W, *et al*. Coverless information hiding algorithm based on image classification[J]. Journal of Hunan University (Natural Sciences), 2019, 46(12): 25-32.(In Chinese)

[7] 吴俊锜,翟黎明,王丽娜,等. 基于多尺度滤波器的空域图像隐写增强算法[J]. 计算机研究与发展,2020,57(11):2251-2259.
WU J Q, ZHAI L M, WANG L N, *et al*. Enhancing spatial steganographic algorithm based on multi-scale filters[J]. Journal of Computer Research and Development,2020,57(11):2251-2259.(In Chinese)

[8] 刘慧超,王志君,梁利平. 基于纹理特性的能量差调制视频水印算法[J]. 湖南大学学报(自然科学版),2020,47(10):116-123.
LIU H C, WANG Z J, LIANG L P. Energy difference modulated video watermarking algorithm based on texture feature[J]. Journal of Hunan University (Natural Sciences),2020,47(10):116-123.(In Chinese)

[9] 付章杰,王帆,孙星明,等. 基于深度学习的图像隐写方法研究[J]. 计算机学报,2020,43(9):1656-1672.
FU Z J, WANG F, SUN X M, *et al*. Research on steganography of digital images based on deep learning[J]. Chinese Journal of Computers,2020,43(9):1656-1672.(In Chinese)

[10] CHEN W H, LIN L, WU M, *et al*. Tackling android stego apps in the wild[C]//2018 Asia-Pacific Signal and Information Processing

Association Annual Summit and Conference (APSIPA ASC). Honolulu, HI, USA: IEEE, 2018: 1564 - 1573.

[11] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, *et al*. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[12] FRIDRICH J, GOLJAN M, LISONEK P, *et al*. Writing on wet paper[J]. IEEE Transactions on Signal Processing, 2005, 53(10): 3923-3935.

[13] PEVNÝ T, FILLER T, BAS P. Using high-dimensional image models to perform highly undetectable steganography[C]//Information Hiding. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010: 161-177.

[14] HOLUB V, FRIDRICH J. Digital image steganography using universal distortion[C]//Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security. New York, NY, USA: Association for Computing Machinery, 2013: 59-68.

[15] FILLER T, JUDAS J, FRIDRICH J. Minimizing additive distortion in steganography using syndrome-trellis codes[J]. IEEE Transactions on Information Forensics and Security, 2011, 6(3): 920-935.

[16] 张政馗,庞为光,谢文静,等. 面向实时应用的深度学习研究综述[J]. 软件学报,2020,31(9):2654-2677.
ZHANG Z K, PANG W G, XIE W J, *et al*. Deep learning for real-time applications: a survey[J]. Journal of Software, 2020, 31(9): 2654-2677. (In Chinese)

[17] 刘勇,李杰,张建林,等. 基于深度学习的二维人体姿态估计研究进展[J]. 计算机工程,2021,47(3):1-16.
LIU Y, LI J, ZHANG J L, *et al*. Research progress of two-dimensional human pose estimation based on deep learning[J]. Computer Engineering, 2021, 47(3): 1-16. (In Chinese)

[18] 谢修娟,顾兵. 基于深度学习的抗菌药物耐药性分析研究[J]. 湖南大学学报(自然科学版),2021,48(10):113-120.
XIE X J, GU B. Research on antimicrobial resistance analysis based on deep learning[J]. Journal of Hunan University (Natural Sciences), 2021, 48(10): 113-120. (In Chinese)

[19] HAYES J, DANEZIS G. Generating steganographic images via adversarial training[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc, 2017: 1951 - 1960.

[20] ZHU J R, KAPLAN R, JOHNSON J, *et al*. HiDDeN: hiding data with deep networks[C]//Computer Vision-ECCV 2018. Munich, Germany: Springer International Publishing, 2018: 682-697.

[21] BERNARD S, BAS P, KLEIN J, *et al*. Explicit optimization of min max steganographic game[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 812-823.

[22] SU A T, MA S, ZHAO X F. Fast and secure steganography based on J-UNIWARD[J]. IEEE Signal Processing Letters, 2020, 27: 221-225.

[23] 沈军,廖鑫,秦拯,等. 基于卷积神经网络的低嵌入率空域隐写分析[J]. 软件学报,2021,32(9):2901-2915.
SHEN J, LIAO X, QIN Z, *et al*. Spatial steganalysis of low embedding rate based on convolutional neural network[J]. Journal of Software, 2021, 32(9): 2901-2915. (In Chinese)

[24] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Proceedings of Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer International Publishing, 2015: 234-241.

[25] YE J, NI J Q, YI Y. Deep learning hierarchical representations for image steganalysis[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(11): 2545-2557.

[26] TANCIK M, MILDENHALL B, NG R. StegaStamp: invisible hyperlinks in physical photographs[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 2114-2123.

[27] AHMADI M, NOROUZI A, KARIMI N, *et al*. ReDMark: Framework for residual diffusion watermarking based on deep networks[J]. Expert Systems with Applications, 2020, 146: 113157.