对抗性增强的图像块位平面拆分缩略图保留加密

李名[a,b],崔清晨[a],王曦[a],张静[a],李文泽[a,b]

(1.河南师范大学 a.计算机与信息工程学院;b.河南省教育人工智能与个性化学习重点实验室,河南 新乡 453007)

摘 要:随着个人图像数量日益增加,云服务开始在图像存储方面发挥重要作用.然而,将图像上传到云端将会面临隐私威胁.传统的加密方案对图像进行简单加密就可以保护图像隐私,但它牺牲了图像内容的可用性.近年来,缩略图保留加密(thumbnail preserving encryption,TPE)被提出,通过加密后保持缩略图不变使云中图像在不被非法第三方肉眼识别的同时能对用户具有可用性.但是现有的 TPE 方案没有考虑到机器学习对图像的隐私威胁.基于此,提出了一种新的 TPE 方案,该方案在加密过程中从对抗深度学习模型识别图像这一全新的角度出发,提高了保护图像隐私信息的能力.实验表明,所提出的方案能够在图像对用户具有可用性的同时抵抗人类肉眼和深度神经网络识别图像.

关键词:图像加密;对抗性;可用性;隐私性

中图分类号:TP309.2 **文献标志码**:A **文章编号**:1000-2367(2025)01-0092-08

随着互联网技术的不断发展,云计算的应用对人们来说变得越来越重要[1],用户可以利用云空间存储文件.智能手机和摄影设备的普及方便了人们随手拍照记录自己的日常.照片的数量日益增多.这些记录日常生活的图像往往会在不经意间透露出人们的隐私信息,比如性别、工作、家庭住址等[2-4].人们通常有意识地保护自己的图像隐私,这在本地设备上非常容易,只需阻止他人浏览自己的个人图像.但随着照片数量的增多,而本地设备资源有限,为了方便图像的存储和使用,越来越多的人将他们的图像上传到云中.用户可以在云中浏览下载自己的图像,无需再在本地设备上备份,这样可以节约本地设备的存储空间,也可以有效避免因为更换硬件或硬件损坏导致图像数据丢失的情况出现.将图像存储到云中的好处是显而易见的,但云中图像不再受用户自己的控制,不能阻止非法第三方从自己的相册中窃取信息.比如来自网络上非法窃取用户隐私的黑客[5].用传统方法将图像加密后再上传到云中,可以有效保护用户的隐私[6-8].但是,用传统加密技术加密的图像在外观上无意义,用户不能直接在云平台上浏览、管理自己的图像.这就使得对于一些普通用户来说,即使他们明白将图像加密后再上传到云中能够很好地保护自己的隐私,为了图像在云中的可用性,他们也会选择直接将明文图像上传到云中.因此,云平台上大都是明文图像,无法保护用户的隐私不受侵犯.保护云中图像的隐私性不应以牺牲图像的可用性为代价.

视觉心理学和图像退化结合的研究成果使密文图像同时具有可用性和隐私性成为可能.许多研究表明[9-10],图像识别是一种基于人类视觉输入和先验知识混合的推理过程,合法用户很容易结合先验知识从图像的退化版本识别出图像,而这对于非法第三方来说很困难.2015 年 WRIGHT 等[11]提出了基于先验知识

收稿日期:2023-06-21;**修回日期**:2023-07-09.

基金项目:国家自然科学基金(61602158);河南省高等学校重点科研项目(23A520009).

作者简介:李名(1981—),男,河南新乡人,河南师范大学副教授,博士,研究方向为图像安全、信息隐藏,E-mail:liming@htu.edu.cn.

通信作者:李文泽(1990—),女,河南新乡人,河南师范大学讲师,博士,研究方向为图像安全、信息隐藏,E-mail:liwenze@htu.edu.cn.

引用本文:李名,崔清晨,王曦,等.对抗性增强的图像块位平面拆分缩略图保留加密[J].河南师范大学学报(自然科学版),2025,53(1):92-99.(Li Ming,Cui Qingchen,Wang Xi,et al.Resistance-enhanced image block bit-plane split thumbnail preserving encryption[J].Journal of Henan Normal University(Natural Science Edition),2025,53(1):92-99.DOI:10.16366/j.cnki.1000-2367.2023.06.21.0003.)

的图像加密概念，即缩略图保留加密.生成的密文图像是原始图像的退化版本，图像所有者能够结合先验知识从密文图像保留的粗略视觉信息中识别图像.但是对于没有浏览过原始图像的其他人，由于没有对原始图像的先验知识，无法识别图像.这就使得密文图像既保护了用户的隐私，又方便用户管理、浏览图像.因此，TPE 方案可以很好地处理图像的隐私性和可用性之间的平衡问题，如图 1 所示.然而，WRIGHT 等[11]提出的仅置换用于加密子块，揭示了原始像素的特定值，经研究表明，无法抵抗统计攻击.TAJIK 等[12]在此基础上提出了新的方案框架，该方案先在块内以两个像素为一组替换加密，再对块内像素置换加密，提高了安全性.但该方案需调用大量具有高计算复杂度的伪随机函数，降低了加密效率.ZHAO 等[13]提出了一种以三像素为一组的 TPE 方案，虽然提高了密文图像缩略图的感知质量，但仍需要对块内像素分组，加解密步骤仍然复杂.YANG 等[14]提出块内像素以随机顺序进行替换加密的 TPEIP 方案，简化了加密过程.

大数据时代下机器学习不断发展，云中图像面临新的隐私威胁.经过完备训练的网络模型可以自动识别图像，但是图像自动识别是一把“双刃剑”，更高的图像识别率往往意味着更多的图像隐私被暴露.云中存储的图像会被隐藏于网络中的机器模型自动识别，带来安全隐患.比如网络上的第三方利用深度网络模型非法识别用户图像并从中获取用户的隐私信息[15].隐私安全是一个很重要的问题，研究云中图像的隐私保护是非常需要的.因此，隐私保护技术要既能防止人类肉眼识别图像窃取信息又能对抗神经网络识别图像窃取信息.而现有的 TPE 方案[11-14]虽然能够阻止非法第三方肉眼识别图像但没有考虑对抗机器学习识别图像.

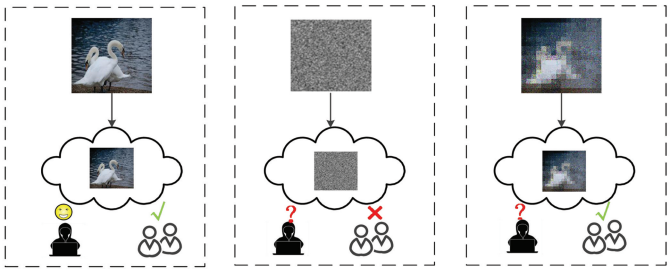

图1　不同类型的图像存储在云平台中

Fig.1　Different types of images are stored on cloud platforms

本文提出一种新的缩略图保留加密方案.该方案以块内所有像素为一个整体进行图像块位平面拆分加密，仅一步操作就可以同时实现替换和置乱.与现有的 TPE 方案相比，该方案打破了分组的限制，像素值改变幅度更大，对抗深度神经网络识别图像的能力更强.对比实验也验证了该方案在对抗深度网络模型识别图像上的有效性.

1　基础知识

1.1　缩略图保留加密

TPE 的核心是在加密前后图像的尺寸和缩略图不变，加密的密文图像呈现出与原始图像的低分辨率版本相同的视觉内容.具体来说，TPE 的加密是先将图像划分为相同维度的子块，对每个块进行像素和保留的加密，块中的像素平均值就是缩略图中该块对应的像素值，缩略图保留加密其实是一种和保留加密.BELLARE 等[16]提出了适用于构建 SPE 的保格式加密方案，步骤如下：

$$\begin{cases} M=(Z_d+1)^n, \\ \vec{\nu}=(\nu_1,\cdots,\nu_n)\in M, \\ S=\sum_{i=1}^{n}\nu_i, \\ Enc_K(T,\vec{\nu})=c, \\ c\in\{\vec{x}\mid\vec{x}=(x_1,\cdots,x_n)\in M, s=\sum_{i=1}^{n}x_i\}, \\ \Phi_{\mathrm{sum}}(c)=\Phi_{\mathrm{sum}}(\vec{\nu})=s, \end{cases} \tag{1}$$

其中，$M=(Z_d+1)^n$ 为消息空间.每个消息都有 n 个单独的元素，这些元素非负，其值不超过 d，一般情况下默认论文中的 $d=255$.$\vec{\nu}$ 是属于 M 中的向量.s 是向量 $\vec{\nu}$ 中元素值的和.Enc_K 是基于随机数的加密算法，c 是

输出的同样属于 M 的密文向量.$\Phi_{\mathrm{sum}}(\vec{v})$ 是求 $\vec{v}$ 中元素值之和的函数.

1.2 主流卷积神经网络结构及识别图像过程

ResNet[17]是一种基于残差结构的卷积神经网络,提出之初是为了解决训练深度卷积神经网络时产生的梯度消失或爆炸的现象,残差结构可以很好地解决这一现象.ResNet 的提出使得卷积神经网络能够实现深度训练,解决了以往网络随着层数加深导致训练效果下降的问题.

VGGNet[18]的网络结构简单、规整且高效.AlexNet 出现之后,很多学者通过改进 AlexNet 的网络结构来提高自己的准确率,主要有两个方向:小卷积核和多尺度.而 VGG 的作者们则选择增加网络深度,并证明了增加网络深度能够在一定程度上影响网络性能.

MobileNet[19]基于流线型架构,使用深度可分离卷积来构建轻量级深度神经网络,用于移动和嵌入式视觉应用.该网络引入了两个简单的全局超参数:宽度乘数和分辨率乘数,可以有效地在延迟和准确性之间进行权衡.这些超参数允许模型构建者根据问题的限制条件为其应用程序选择合适大小的模型.

利用神经网络对隐私图像进行识别的过程:非法第三方首先获得用户在云平台上传的图像,然后通过使用大型自然拍摄图像的数据集训练好的深度网络模型分类器对图像进行识别分类,根据用户图像的主要类别对用户身份等隐私信息进行解密.

2 对抗性增强的图像块位平面拆分加密算法

GEIRHOS 等[20]研究证实深度学习模型在学习过程中存在过度依赖图像纹理特征进行决策的现象.出于对此现象的分析,本文通过对图像位平面置乱让深度学习网络作出错误决策.该方法通过改变像素的值及位置来将图像原有信息打乱,也就是改变图像的纹理特征.一些研究指出深度学习模型只是利用了局部区域的特征聚合进行分类,因此本文为了提升算法对模型的攻击能力,提出分区域位平面置乱,将原始图像划分成更小的次级区域结构进行位平面拆分加密,图 2 是区块以二进制位平面拆分置乱的例子.对原始图像局部特征聚合的干扰而言,这样的方式对深度学习网络的对抗更强.

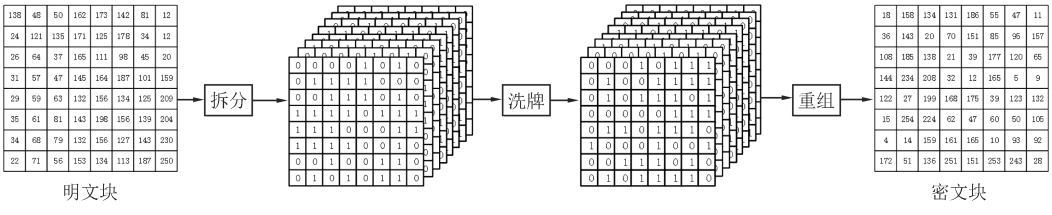

图2 块以二进制进行位平面拆分加密的流程图

Fig.2 Flowchart of bitplane splitting and encryption in binary for block

块位平面拆分保证了块的像素值之和不变,保留了一定程度上的视觉可用性.与此同时,该方法以整个块作为一个整体加密,改变了块中每个像素的灰度值,且改变像素值的范围不受限制.因此,理论上来说,块位平面拆分置乱具有很强的对抗性.为了验证该算法的有效性,对一张花卉图像分别用块位平面拆分置乱、块内像素直接置乱以及块内像素用均值替换这 3 种方法加密,然后对比加密图像的对抗性.表 1 给出了主流的网络模型能将图像正确分类的概率.可以看出这 3 种分块进行加密的操作都对神经网络模型具有较强的攻击效果,块位平面拆分置乱的对抗性更强.其中图像都是以 14×14 块大小进行加密,因为该块大小加密的密文图像在隐藏图像细节信息的同时从人类视觉上仍能看出其大概类别.图 3 展示了原始图像以及分别用块位平面拆分置乱、块内像素直接置乱以及块内像素用均值替换这 3 种方法加密的密文图像的视觉效果.

块位平面拆分加密算法不仅具有对抗性增强的特点,还具有更易实现缩略图保留加密的以下特点.

块和不变:将图像分成大小相等的子块,分别对每个子块进行位平面拆分置乱加密,块中的像素值被改变,但仍能保持块中所有像素的值之和不变.位平面拆分置乱加密可以实现和保留加密.

同时实现置乱和扩散:对每个子块进行位平面拆分加密能够在改变像素位置的同时改变像素的值,置乱和扩散同时进行.置乱和扩散是公认的能保证图像加密安全的手段,一般都是分别进行,会增加被攻击风险.

位平面拆分加密算法置乱和扩散同时进行，能够提高安全性.

表 1　图像被正确分类的概率

Tab. 1　The probability of images being correctly classified

模型	原图	块位平面拆分置乱	块内像素置乱	块均值替换
VGG16[18]	0.998 5	0.002 3	0.003 1	0.012 4
ResNet50[17]	0.999 7	0.000 1	0.000 1	0.003 0
MobileNetV2[19]	0.999 9	0.002 5	0.002 6	0.002 8

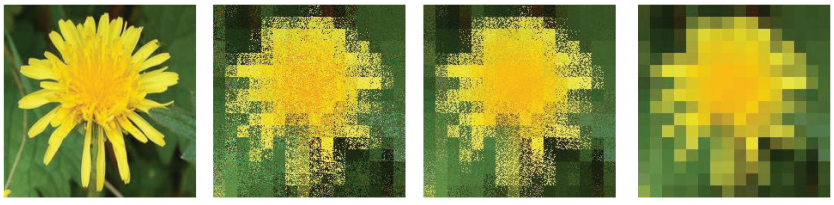

图3　图像的视觉效果

Fig. 3　Visual effects of images

加解密效率高：对图像块进行位平面拆分加密算法可以仅一步实现块和保留替换加密，步骤简单.和现有的一些缩略图保留加密算法相比，位平面拆分加密算法不需要对块内像素分组，不需要调用大量具有高计算复杂度的伪随机函数，图像加密解密效率更高.

3　提出的图像缩略图保留加密方案框架

基于对抗性增强的块位平面拆分加密的技术特点，提出块位平面拆分加密的缩略图保留加密方案.该方案能够在图像具有可用性的同时抵抗人类肉眼和深度神经网络识别图像.在这个方案中，使用以下类型的图像：每张图像都有一个或多个通道组成，例如灰度、RNG、PNG；每张图像都有一组唯一的公共元数据，可以将其视为随机数，像现有的 TPE 方案一样充当标识符(即 T)；图像的每个通道都是由像素所组成的二维矩阵，其中每个像素都属于 Z_{d+1}(在下文中如果没有特殊说明，$d=255$).

3.1　图像加密

彩色图像通常由多个通道组成，该方案单独处理每个通道图像，将每个通道矩阵分割成大小为 $b\times b$ 的缩略图块，以块中所有像素为一个整体进行加密.TPE 方案的核心在于每个缩略图块中的像素值之和在加密前后保持不变，该方案通过保证每个通道中块的像素和不变来保证整个图像中块的像素和不变.图像整体加密方案如算法 1 所示.对图像分成大小相等的缩略图块 $M_b[i,j,k]$，每个块单独加密处理，加密过程可以重复多个轮次，记为 R，加密后仍保持块和不变.

算法 1　图像整体加密算法

输入：K，T，b，R and M

输出：C

1：(m,n,c)=params(M)/ * 得到原始图像的尺寸 * /

2：for $r=1:R$ do

3：　for $k=1:c$ do

4：　　for $i=1:m/b$ do

5：　　or $j=1:n/b$ do

6：　　$T'=T\parallel r\parallel k\parallel i\parallel j$

7：　　$C_b[i,j,k]=SPEnc_K(T',M_b[i,j,k])$/ * 位平面置乱的和保留加密 * /

8：　　end for

9：　end for

10：　end for

11：　end for

12：return C

对缩略图块进行位平面拆分，为了提高加密方案的安全性，不同的块进行位平面拆分采用的进制不同，每个块所采用的进制与随机的前一块建立联系得到.该方案对不同的块随机以二进制、四进制、八进制、十六进制之间的任意一个进制做位平面拆分，因为这 4 个进制可以让像素值在拆分置乱重组后仍在[0，255]的范围内.通道图像中每个块进行位平面拆分所采用进制的生成方法如式(2)所示.通过引入混沌系统生成一个

随机数序列,前一个块进行位平面拆分所采用的进制加上随机数取模得出一个整数,根据整数所在的范围选择该块进行位平面拆分采用的进制.可以多轮扩散,通过改变一个块所采用的进制可以影响尽可能多的块所采用的进制.

$$x_{i+1}=(k_{i+1}+t_i)\bmod 16,$$
$$t_{i+1}=\begin{cases}2, 0\le x_{i+1}<4,\\ 4, 4\le x_{i+1}<8,\\ 8, 8\le x_{i+1}<12,\\ 16, 12\le x_{i+1}<16,\end{cases} \tag{2}$$

其中,k_{i+1} 是引入混沌系统生成的随机序列中的第 $i+1$ 个随机数,t_i 是上一个块进行位平面拆分采用的进制,t_{i+1} 是通道图像中第 $i+1$ 个块进行位平面拆分采用的进制.

对块以不同进制做位平面拆分后分别对不同的位平面进行置乱,在改变块中像素值的同时保持块和不变.首先将块按其进制做位平面拆分,然后对每个位平面进行洗牌操作,最后将置乱后的位平面重组得到加密后的图像块.如式(3)所示,块中像素值先以 t 进制做位平面拆分,对位平面置乱后再以 t 进制重组得到密文像素值.

$$X=x_n x_{n-1}\cdots x_1, X'=x_1'+x_2'\times t+\cdots+x_n'\times t^{n-1}, \tag{3}$$

其中 X 是明文像素值,X'是密文像素值,t 是该块进行位平面拆分采用的进制.

3.2 图像解密

在解密过程中,不同的图像需要不同的密钥,这些密钥与加密过程中使用的密钥相同.解密过程与加密过程类似,单独处理每个通道图像并对其分块,然后对每个块进行解密操作.解密时,利用密钥中的信息得到特定密钥,得到与加密过程中使用的相同随机数序列.加随机数取模得到每个块进行位平面拆分所采用的进制.

4 实验分析

在本节中,分别从图像质量、对抗深度学习模型识别图像等方面验证该方案的性能,充分展示了所提出方案的优势.

数据集:ImageNet 数据集和 Oxford-102flower 数据集都是收集的自然拍摄图像,因此选择这两个数据集测试该方案的性能.从 ImageNet 数据集中随机选取 10 类,其中每类选择 100 张共 1 000 张图像.Oxford-102flower 数据集共 102 类,从每类随机选择 10 张共 1 020 张图像.同时,将选取的所有图像的分辨率调整为 224×224.

4.1 图像质量

该方案加密后的密文图像保留了和原始图像完全相同的缩略图,是精确的 TPE.一般来说,密文图像的缩略图与原始图像的缩略图越相似表明方案越优,这意味着本文的方案达到了最优的加密状态.同时该方案可以完全解密出原始图像,解密后的图像效果也达到了理论最优.因此本文提出的方案在图像质量方面不低于其他现有的方案.对图像加密的视觉效果如图 4 所示,从左到右分别为 224×224 的原始图像、块大小分别为 7×7、14 × 14、28×28 的加密图像,可以看出用户可以通过调整块大小来调整图像隐私性和可用性之间的平衡.

4.2 对抗深度学习模型识别图像的性能

本节对调整后的 ImageNet 数据集和 Oxford-102flower 数据集进行缩略图保留加密,测试图像以不同块大小进行缩略图保留加密后的密文图像对抗分类器的性能.为了证明本文方法的普遍有效性,选择 3 个深度神经网络模型进行研究,并以攻击模型的成功率作为标准与 TPEIP[14] 方案进行对比,因为 TPEIP 方案同本文的方案一样都是仅一步的精确 TPE,因此选择其作为对比方案.

图像以 7×7、14×14 的块大小进行 TPE 的密文图像隐藏了图像中的细节信息,但从人类视觉上仍能看出其大概类别.因此主要分析这两个块大小对以上 3 种不同类型的网络模型进行缩略图保留加密攻击的结

果.表 2 给出了网络模型对该方案及 TPEIP 方案分别以 7×7、14×14 的块大小进行一轮加密的图像分类准确率，其中原始准确率是指网络模型对原始测试图像分类的准确性.可以看出这两个 TPE 方案都对神经网络模型具有较强的攻击效果，并且本文的方案对抗网络模型识别图像的能力优于 TPEIP.

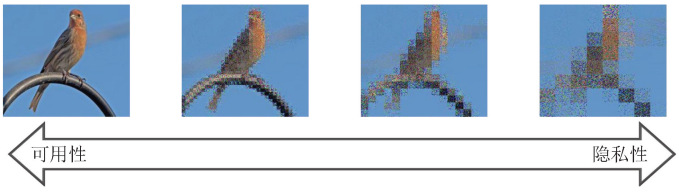

图4　加密图像的效果

Fig. 4　The effect of encrypting images

表 2　该方案与 TPEIP 方案对抗神经网络模型识别图像的性能比较

Tab. 2　Performance comparison of this scheme with the TPEIP scheme against neural network models for image recognition

数据集	加密块大小	方案	MobileNetV2[19]/%	ResNet50[17]/%	VGG16[18]/%
ImageNet 数据集		原始准确率	96.1	97.6	96.5
	7×7	该方案	41.8	40.1	18.2
		TPEIP	43.3	48.6	18.4
	14×14	该方案	27.5	19.8	14.3
		TPEIP	28.1	32.0	14.5
Oxford-102flower 数据集		原始准确率	96.86	96.86	87.94
	7×7	该方案	12.98	13.24	9.61
		TPEIP	13.43	17.25	5.88
	14×14	该方案	4.51	2.35	2.84
		TPEIP	7.16	7.06	3.53

4.3　尺寸扩展率

所提出的方案没有产生额外的数据，密文图像是和原始图像具有相同维度的二维矩阵，所以从理论上讲，它们的大小应该相同.然而，由于 JPG 等格式的图像使用了图像压缩算法，密文图像的信息冗余度小于原始图像，因此，与原始图像相比，密文图像的压缩率低，从而会出现一定的密文尺寸膨胀.对 ImageNet 数据集中随机选取的 1 000 张图像分别进行该方案和 TPEIP 方案加密.不同块大小加密的图像尺寸扩展的结果如表 3 所示.结果表明，该方案加密后图像数据集的尺寸扩展率小于 2.3，比 TPEIP 方案[14]低.

表 3　该方案与 TPEIP 方案密文尺寸扩展率的比较

Tab. 3　Comparison of ciphertext size expansion rate between this scheme and TPEIP scheme

加密块大小	7×7	14×14	28×28	56×56
该方案	1.65	1.81	1.99	2.26
TPEIP	2.33	2.37	2.43	2.51

4.4　像素相关性分析

由于自然拍摄的图像具有较强的空域相关性，图像中的像素通常会泄露其周围像素的信息.攻击者通常利用这一特性预测周围像素的灰度值，从而恢复明文图像.所以必须破坏相邻像素之间的相关性以防止统计攻击.表 4 展示了大小为 256×256 的 Lena 原始图像和该方案加密的密文图像中相邻像素的相关性.可以看出，加密前图像在水平、垂直、对角线 3 个方向上的相关性系数都在 0.9 以上，具有很强的空域相关性.加密图像的相关性系数在 3 个方向上都比原始图像的相关性系数降低了 0.5 以上，所以该方案能够在一定程度上抵抗统计攻击.

4.5　安全性分析

安全性是指非法的第三方无法从粗略的密文图像中学习到关于明文图像的任何确切信息.为了分析本文方案的安全性，将加密方案建模为马尔可夫链，并将其安全性与该马尔可夫链的混合时间相关联.马尔可

夫链的混合时间是在马尔可夫链状态上达到 ε-close 平稳分布所需的最小轮数[21].首先,介绍马尔可夫链的几种定义.

定义 1 有限马尔可夫链是基于转移概率矩阵 $\boldsymbol{P}$ 在有限集 Ω 中移动元素的过程.$\boldsymbol{P}$ 是随机的,所有元素都是非负的,每一行元素的和为 1.

定义 2 假设在有限集 Ω 上有一个分布 π,满足 $\pi=\pi\boldsymbol{P}$,则称 π 是马尔可夫链的平稳分布[21].经过足够的回合,马尔可夫链状态的分布接近平稳分布.

将加密方案建模为一个马尔可夫链.以缩略图块中所有像素为一组,对于任意和为 s 的像素组,Z 表示和为 S 的所有像素组的集合.每个像素组都是马尔可夫链的一个状态,马尔可夫链的转移矩阵代表从一个状态转移到另一个状态的概率,也就是由一个像素组加密为另一个像素组的概率,这个概率由位平面拆分置乱算法的概率确定.马尔可夫链以理想的方式对概率进行建模.马尔可夫链的混合时间是当马尔可夫链达到平稳状态所需要的时间,在本文加密方案中就是加密达到一定安全要求所需要的轮数,根据式(4)求出加密方案所需要的轮数[12]:

$$t_{\mathrm{mix}}(\varepsilon)=\lceil\frac{2(\ln\varepsilon-\ln(|\Omega|-1))}{\ln\lambda_{*}}\rceil. \tag{4}$$

其中,$|\Omega|$ 是马尔可夫链状态转移矩阵的状态数目,λ_{*} 是马尔可夫链状态转移矩阵与其转置矩阵乘积的第二大特征值,符号$\lceil x\rceil$表示不小于 x 的最小整数.根据文献[12],敌手无法区分 Enc_K 和随机 Φ 保持函数,因此该加密方案满足 NR-安全性.

表 4 不同方向相邻像素的相关系数

Tab. 4 Correlation coefficient of adjacent pixels in different directions

块大小	通道	水平	垂直	对角
原始图像	R	0.97	0.95	0.93
	G	0.97	0.94	0.92
	B	0.95	0.92	0.90
加密图像	R	0.39	0.37	0.37
	G	0.34	0.33	0.33
	B	0.22	0.21	0.23

5 总 结

云平台用户将图像上传到云中,希望云中图像既能不泄露隐私信息又能对用户具有可用性.本文提出了一个新的 TPE 方案,该方案加密的图像可以在保护图像隐私和对用户具有可用性之间实现很好的平衡,同时用户可以在需要时完全解密出原始图像.与已有研究相比,新方案不用对块内像素分组,加密流程简单.同时,经过实验表明,在自然拍摄的图像数据集上该方案相比于对比方案对抗深度神经网络识别图像的能力更强,可以说明在真实用户情况下该方案的对抗性也更强,能够更好地保护用户图像信息安全.

参 考 文 献

[1] FOSTER I,ZHAO Y,RAICU I,et al.Cloud computing and grid computing 360-degree compared[C]//2008 Grid Computing Environments Workshop.[S.l.]:IEEE,2008.

[2] YU J,ZHANG B P,KUANG Z Z,et al.iPrivacy:image privacy protection by identifying sensitive objects via deep multi-task learning[J].IEEE Transactions on Information Forensics and Security,2017,12(5):1005-1016.

[3] ZHANG L,JUNG T,LIU K B,et al.PIC:enable large-scale privacy preserving content-based image search on cloud[J].IEEE Transactions on Parallel and Distributed Systems,2017,28(11):3258-3271.

[4] IIDA K,KIYA H.Privacy-preserving content-based image retrieval using compressible encrypted images[J].IEEE Access,2020,8:200038-200050.

[5] SOLANKE VIKAS S,PAWAN K,KULKARNI GURUDATT A,et al.Mobile cloud computing:Security threats[C]//2014 International Conference on Electronics and Communication Systems(ICECS).Coimbatore:IEEE,2014.

[6] LAN R S，HE J W，WANG S H，et al.Integrated chaotic systems for image encryption[J].Signal Processing，2018，147：133-145.
[7] CHUMAN T，SIRICHOTEDUMRONG W，KIYA H.Encryption-then-compression systems using grayscale-based image encryption for JPEG images[J].IEEE Transactions on Information Forensics and Security，2019，14(6)：1515-1525.
[8] YE G D，PAN C，DONG Y X，et al.Image encryption and hiding algorithm based on compressive sensing and random numbers insertion [J].Signal Processing，2020，172：107563.
[9] GREGORY R L.Knowledge in perception and illusion[J].Philosophical Transactions of the Royal Society of London Series B，Biological Sciences，1997，352(1358)：1121-1127.
[10] KINJO H，SNODGRASS J G.Does the generation effect occur for pictures?[J].The American Journal of Psychology，2000，113(1)：95-121.
[11] WRIGHT C V，FENG W C，LIU F.Thumbnail-preserving encryption for JPEG[C]//Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security.[S.l.]：ACM，2015.
[12] TAJIK K，GUNASEKARAN A，DUTTA R，et al.Balancing image privacy and usability with thumbnail-preserving encryption[C]//Proceedings 2019 Network and Distributed System Security Symposium.San Diego：Internet Society，2019.
[13] ZHAO R Y，ZHANG Y S，XIAO X L，et al.TPE2：three-pixel exact thumbnail-preserving image encryption[J].Signal Processing，2021，183：108019.
[14] YANG C H，WENG C Y，YANG Y Z.TPEIP：Thumbnail preserving encryption based on sum preserving for image privacy[J].Journal of Information Security and Applications，2022，70：103352.
[15] SAEIDIAN S，CERVIA G，OECHTERING T J，et al.Quantifying membership privacy via information leakage[J].IEEE Transactions on Information Forensics and Security，2021，16：3096-3108.
[16] BELLARE M，RISTENPART T，ROGAWAY P，et al.Format-preserving encryption[M].Heidelberg：Springer，2009：295-312.
[17] HE K M，ZHANG X Y，REN S Q，et al.Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR).[S.l.]：IEEE，2016：770-778.
[18] SIMONYAN K，ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[C].3rd International Conference on Learning Representations.[S.l.：s.n.]，2015.
[19] SANDLER M，HOWARD A，ZHU M L，et al.MobileNetV2：inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.[S.l.]：IEEE，2018.
[20] GEIRHOS R，RUBISCH P，MICHAELIS C，et al.ImageNet-trained CNNs are biased towards texture；increasing shape bias improves accuracy and robustness[EB/OL].[2023-06-10].http://arxiv.org/abs/1811.12231v3.
[21] LEVIN D A，PERES Y，WILMER E L.Markov chains and mixing times[M].2nd ed.Providence：AMS，2017.

Resistance-enhanced image block bit-plane split thumbnail preserving encryption

Li Ming[a,b]，Cui Qingchen[a]，Wang Xi[a]，Zhang Jing[a]，Li Wenze[a,b]

(a. College of Computer and Information Engineering；b. Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province，Henan Normal University，Xinxiang 453007，China)

Abstract：With the increasing number of personal images，cloud services are beginning to play an important role in image storage. However，uploading images to the cloud leads to privacy threats. Traditional encryption schemes that simply encrypt images can protect image privacy，but they sacrifice the usability of the image content. In recent years，thumbnail preserving encryption(TPE) has been proposed to make images available to users in the cloud without being identified by illegal third parties with the unchanged thumbnails after encryption. However，existing TPE schemes do not take account of the privacy threat to images from machine learning. We propose a new TPE scheme that improves the ability of the algorithm to protect image privacy information from a new perspective by combating deep learning models in the encryption process. Experiments show that the proposed scheme is able to resist image recognition by human eyes and deep neural networks while the image is usable.

Keywords：image encryption；adversarial；usability；privacy

[责任编校　陈留院　杨浦]