文章编号:1000-2367(2020)01-0033-05 DOI:10.16366/j.cnki.1000-2367.2020.01.006

基于生成对抗网络的无载体信息隐藏

段新涛,李宝霞,贾凯,郭玳豆

(河南师范大学 计算机与信息工程学院,河南 新乡 453007)

摘 要:提出了一种基于生成对抗网络的无载体信息隐藏方法.发送方用伪装图像和 Improved Wasserstein GAN 生成一张与秘密图像相同的图像,发送方发送伪装图像,接收方收到伪装图像后用生成器生成和秘密图像视觉上相同的图像.与现有的信息隐藏方法相比,该方法不需要修改载体,能够有效地抵抗隐写分析工具的检测,能够达到秘密信息安全传输的目的.

关键词:Improved Wasserstein GAN;无载体信息隐藏;信息安全

中图分类号:TP309.7 **文献标志码**:A

目前大多数信息隐藏技术是通过对载体数据(数字图像、视频和音频等)的修改将秘密信息嵌入到载体中,并将其隐藏起来作为隐藏的秘密信息,而数字图像包含大量信息且使用最为广泛,常被当作理想的信息隐藏载体[1-2].图像隐写技术主要分为空间域和变换域两类:空间域隐藏技术是通过直接改变隐藏数据中图像像素值的某些位,来实现信息的隐藏[3].如:最低有效位(Least Significant Bit,LSB)信息隐藏,该算法在最低有效位算法(LSB)和选定最低有效位(SLB)算法之间达到平衡,从而实现了安全性和图像质量之间的平衡,但 LSB 信息隐藏方法缺乏健壮性,隐藏的数据可能会丢失,隐藏数据很容易被破坏,隐秘信息容易被发现[4-5].像素差分(Pixel-value differencing,PVD)信息隐藏是不可逆数据隐藏方法,该方法将秘密信息嵌入到被划分为两个连续像素的非重叠块儿的灰度载体图像中[6].变换域信息隐藏,它们是一种隐藏图像中信息的更复杂方式,是指在图像上使用各种算法和变换来隐藏秘密信息[7].基于离散傅立叶变换(Discrete Fourier Transform,DFT)的隐写方法,图像的 DFT 的大小被舍入并用位平面表示,秘密信息被嵌入到位平面中[8].基于 DCT 的数据隐藏方法,该方法通过颜色量化、颜色排序、数据隐藏等步骤实现图像的隐写[9].基于 DCT 和 LSB 的图像隐写方法,该方法将隐秘图像嵌入到载体图像中[10].这些信息隐藏方法都是通过对载体按照一定的规则进行修改来嵌入秘密信息,不可避免地把修改痕迹留在载体上,因此以上方法很难抵抗各类隐写算法的检测[11-16].

本文提出一种基于生成对抗网络 Improved Wasserstein GAN 的信息隐藏.发送方用秘密图像和 Improved Wasserstein GAN 生成模型生成一张与秘密信息无关的自然图像,发送伪装图像,接收方用接收到的图像和生成器生成和秘密图像视觉上相同的图像.与以上的信息隐藏方法相比,此方法没有对载体进行任何的修改,实验表明本文的方法能够有效抵抗隐写分析工具的检测.本文所做的主要工作有:(a)使用了文献[17]中提出的新的 Lipschitz 连续性限制手法——梯度惩罚,解决了 Wasserstein GAN 训练梯度消失梯度爆炸的问题;(b)比标准 Wasserstein GAN 拥有更快的收敛速度,并能生成更高质量的样本;(c)提供稳定的 GAN 训练方式,调参简单,成功训练多种针对图片生成模型的 GAN 架构.

1 模型简介

一个 GAN 主要包含两个独立的神经网络:生成网络和判别网络,生成网络就像造假币团队,判别网络

收稿日期:2018-10-16;**修回日期**:2019-08-09.

基金项目:河南省高等学校重点科研项目(19B510005;20B413004;16A520058)

作者简介:段新涛(1972-),男,河南新乡人,河南师范大学副教授,博士,研究方向为信息隐藏,信息安全.

通信作者:李宝霞,E-mail:2939262407@qq.com.

就像警察试图检测假币,双发都互相改变自己的方法直到伪造品与真品无法区分[18].GAN 训练目标如下:

$$\min_G \max_D V(D,G)=E_{x\sim p_{data}}(x)[\lg D(x)]+E_z \sim P_z(z)[\lg(1-\lg D(G(z)))], \tag{1}$$

其中 x 表示真是图片,Z 表示输入生成网络的噪声,而 $G(Z)$表示生成网络生成的图片.判别器损失最小化:

$$-E_{z-pr}[\lg D(x)]-E_{x-p_g}[1-\lg D(x)]. \tag{2}$$

Goodfellow 最初提出的生成网络损失(最小化):

$$E_{x-p_g}[1-\lg D(x)]. \tag{3}$$

训练过程中先固定生成网络,训练判别网络达到最优,然后训练生成网络,利用 SGD 训练判别网络达到最优解为:

$$D^*(x)=\frac{p_r(x)}{p_r(x)+p_g(x)}, \tag{4}$$

最终变换形式:

$$KL(p_g \parallel p_r)-2JS(p_r \parallel p_g), \tag{5}$$

最小化 p_r 和 p_g 之间的 JS 散度,但是 p_r 和 p_g 会有忽略的重叠,所以无论它们相距多远 JS 散度都是常数 lg 2,最终导致生成器的梯度(近似)为 0,会导致梯度消失.后来改进的生成器损失:

$$E_{x-p_g}[-\lg D(x)]. \tag{6}$$

最小化目标分析:

$$KL(p_1 \parallel p_2)=E_{x\sim p_1}\lg\frac{p_1}{p_2}, \tag{7}$$

$$JS(p_1 \parallel p_2)=\frac{1}{2}KL(P_1 \parallel \frac{P_1+P_2}{2})+\frac{1}{2}KL(\frac{P_1+P_2}{2}), \tag{8}$$

最小化生成分布与真实分布的 KL 散度,却又要最大化两者的 JS 散度,在数值上则会导致梯度不稳定.Wasserstein 距离:

$$W(p_r,p_g)=\frac{1}{K}\inf_{r\sim\prod(p_r,p_g)}E_{(x,y)-r}[\parallel x-y \parallel]. \tag{10}$$

Improved Wasserstein GAN 的对偶问题:

$$W(p_r,p_g)=\frac{1}{K}\sup_{\parallel f \parallel L\leq K}E_{x\sim p_r}[f(x)]-E_{x-p_r}[f(x)], \tag{11}$$

要求函数 f 的导函数绝对值不超过 K 的条件下,对所有可能满足条件的 f 取到(11)式的上界,然后再除以 K.用该距离做 GAN 的损失函数,可得生成网络损失函数:

$$-E_{x\sim p_r}[f(x)]. \tag{12}$$

判别网络损失函数:

$$E_{x\sim p_g}[f(x)]-E_{x\sim p_r}[f(x)] \tag{13}$$

可以表示训练过程中,其数值越小表示真实分布与生成分布的 Wasserstein 距离越小,GAN 训练得越好.Wasserstein 距离相对 KL 散度与 JS 散度具有优越的平滑特性,可以解决梯度消失问题.在最优判别网络下优化生成网络使得 Wasserstein 距离缩小,能有效拉近生成分布与真实分布.Wasserstein GAN 既解决了训练不稳定的问题,也提供了一个可靠的训练进程指标.

Wasserstein GAN 对原始 GAN 进行了改进,提出了一种替代 Wasserstein GAN 判别网络中权重剪枝的方法,即具有梯度惩罚的 Wasserstein GAN,从而避免训练不稳定的情况[17].但是在 Wasserstein GAN 中,如果将权重剪切到一定范围内.会发现大部分权重都在极端,也就是大部分参数会走极端,要么取最大值要么取最小值,无法发挥深度神经网络的泛化能力,剪切范围太小导致梯度消失,而剪切范围太大,梯度变大一点点,多层以后梯度就会爆炸.为了解决这个问题,Improved Wasserstein GAN 引入了 Lipschitz 连续,其实就是在一个连续函数 f 上额外加了一个限制,要求存在一个常数 $K>=0$ 使得定义域内的任意两个元素 x_1 和 x_2 都满足:

$$|f(x_1)-f(x_2)|\leq K|x_1-x_2|, \tag{14}$$

此时称函数 f 的 Lipschitz 常数为 K.Wasserstein GAN 的目标函数公式:

$$W(p_r,p_g)=\frac{1}{K}\sup_{\|f\|_L\le K}E_{x\sim p_r}[f(x)]-E_{x-p_g}[f(x)]. \tag{15}$$

是由上述 Wasserstein 距离公式产生,可以用一组参数 W 来定义一系列可能的函数 f_w,此时求解(15)式可以近似变成求解如下形式:

$$K\cdot W(p_r,p_g)\approx\max_{W:|f_w|_L\le K}E_{x\sim p_r}[f_w(x)]-E_{x\sim p_g}[f_w(x)]. \tag{16}$$

只需显示神经网络 f_w 的所有参数 f_i 的值不超过某个范围(不能为正无穷),Lipschitz 连续条件要求判别网络的梯度不超过 K,可以额外的设置一个损失项来体现这一点:

$$\mathrm{Re}\,LU[\|\nabla_x D(x)\|_p-K], \tag{17}$$

判别网络在训练好以后,梯度就会在 K 附近,通过这一点可以把上面的损失改成要求梯度越接近 K 越好:

$$[\|\nabla_x D(x)\|_P-K]^2, \tag{18}$$

如果把 K 设为 1,再跟 Wasserstein GAN 原来的判别网络损失加权合并,就得到新的判别网络损失:

$$L(D)=-E_{x\sim p_r}[D(x)]+E_{x\sim p_g}[D(x)]+\lambda E_{x\sim\xi}[\|\nabla_x D(x)\|p-1]^2. \tag{19}$$

Improved Wasserstein GAN 模型解决了 Wasserstein GAN 收敛速度慢的问题,而且生成样本质量高,调参简单[17].GAN 可用于生成手写字体,Wasserstein GAN 解决了 GAN 模式不稳定的问题,文献[19]又提出 Wasserstein GAN 用于无载体信息隐藏技术,并解决了 GAN 模型崩塌问题.因此,本文将 Improved Wasserstein GAN 模型用于信息隐藏技术.

2　实验环境、数据集及步骤

在 tensorflow1.1.0,GPU1080 环境下实现该算法,随机收集 5 000 张图像并处理成大小为 256×256 大小的灰度图像进行实验.基于 Improved Wasserstein GAN 模型的信息隐藏算法实验流程如图 1 所示,具体实现步骤为:(a)对伪装图像、秘密图像进行预处理,构建图像数据库;(b)构建 Improved Wasserstein GAN 生成对抗网络模型,初始化参数;(c)利用图像数据库训练生成对抗网络的模型;(d)将训练好的生成器参数保存,构建生成模型数据库;(e)传输伪装图像,接收方接收伪装图像,用生成器生成和秘密图像视觉上相同的图像.

图 1　实验流程图

Fig.1　Experimental flowchart

3　实验及结果分析

3.1　信息隐藏过程

由 Improved Wasserstein GAN 模型生成手写字体,输入的是随机噪声,生成与原始图形无关的手写字体图像.因此,想到用一张图像输入到模型中,生成一张和隐秘图像视觉上相同的自然图像,从而传输伪装图像,再用生成图像和生成网络生成秘密图像,来达到与传输伪装图像和传输隐秘图像一样的效果,下面以 Lena,Baboon,Cameraman 和 Peppers 来进行试验结果展示.隐藏过程如图 2 所示.

3.2　秘密图像恢复过程

接收方用伪装图像和相应的生成网络,生成和秘密图像视觉上相同的图像.传输的是与秘密图像无关的自然图像,在传输过程中很难被攻击者发现.

从图 3 可以观察到,生成的图像和秘密图像视觉上相同,在传输过程中发送方只需传输伪装图像,接收方把伪装图像输入到生成网络,生成和秘密图像相同的图像即可实现和传输秘密图像相同的效果.

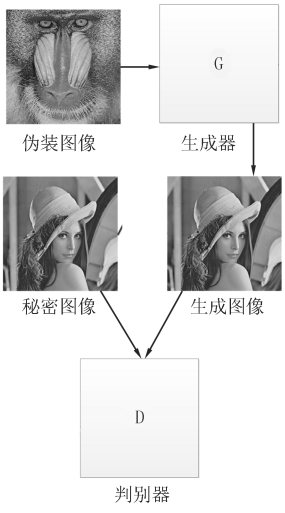


图 2 信息隐藏过程

Fig. 2 Information hiding process

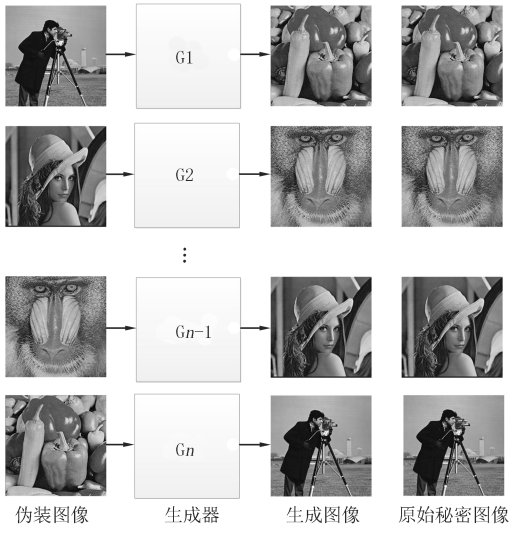
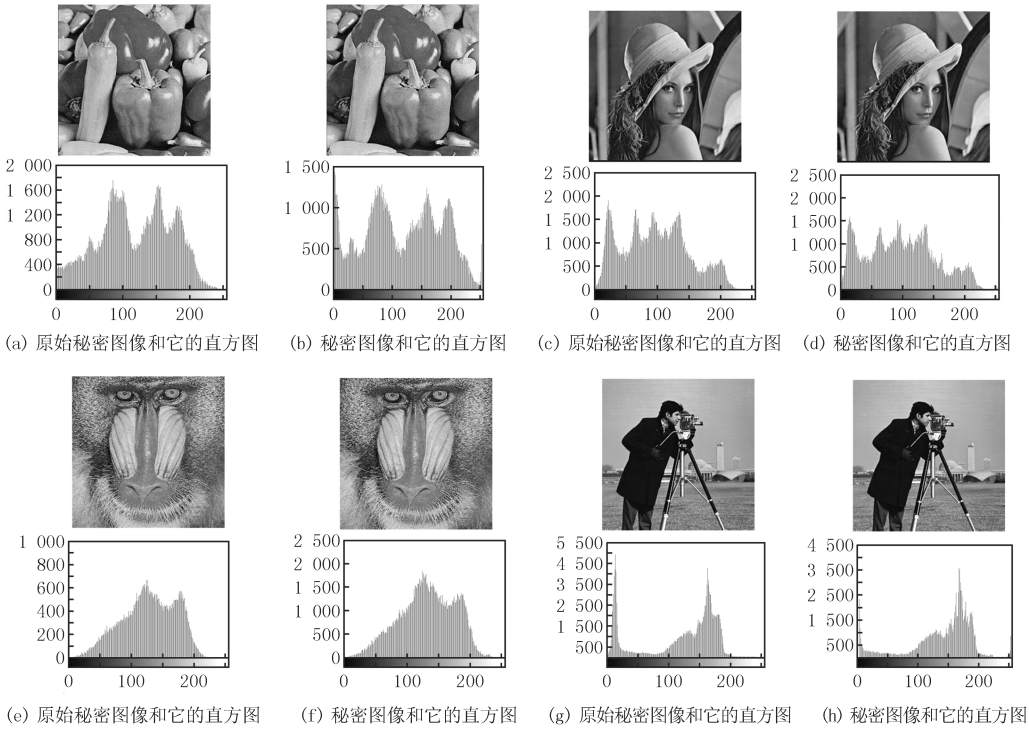


图 3 接收方恢复秘密图像过程

Fig. 3 The receiver recovers the secret image

3.3 实验结果分析

除了对生成图像和原始秘密图像做了视觉上的对比,为了说明该方法的可行性另外补充了生成图像和原始秘密图像的直方图分析,结果分析如图 4 所示.



(a) 原始秘密图像和它的直方图 (b) 秘密图像和它的直方图 (c) 原始秘密图像和它的直方图 (d) 秘密图像和它的直方图

(e) 原始秘密图像和它的直方图 (f) 秘密图像和它的直方图 (g) 原始秘密图像和它的直方图 (h) 秘密图像和它的直方图

图 4 原始秘密图像和生成图像直方图分析

Fig. 4 Histogram analysis of original secret image and generated image

从原始秘密图像和生成图像的直方图中可以看到直方图形状非常相似,换句话说生成图像可以从视觉上代替原始秘密图像,证明该方法是可行的.

4 结 论

信息安全的传输对于国家和个人而言有重大意义,数字图像又是信息量极大的传输介质.本文提出了基于生成对抗网络的无载体信息隐藏,此方法不需要对载体进行修改,发送方发送伪装图像,接收方就能用伪

装图像和生成器生成和秘密图像视觉上相同的图像,实验结果表明该方法能够实现秘密信息的安全传输.

参 考 文 献

[1] LI B,HE J,HUANG J.A survey on image steganography and steganalysis[J].Department of Computing,2011,2(3):288-289.

[2] 张新鹏,殷赵霞.多媒体信息隐藏技术[J].自然杂志,2017,39(2):87-95.
ZHANG X P,YIN Z X.Multimedia Information Hiding Technology[J].Nature Magazine,2017,39(2):87-95.

[3] 张新鹏,钱振兴,李晟.信息隐藏研究展望[J].应用科学学报,2016(5):475-489.
ZHANG X P,QIAN Z X,LI S.Prospects of Information Hiding Research[J].Journal of Applied Sciences,2016(5):475-489.

[4] OSUNADE O,ADENIYI G I.Enhancing the Least Significant Bit(LSB) Algorithm for Steganography[J].International Journal of Computer Applications,2016,149(3):1-8.

[5] CHAN C K,CHENG L M.Hiding data in images by simple LSB substitution[J].Pattern Recognition,2004,37(3):469-474.

[6] WU D C,TSAI W H.A steganographic method for images by pixel-value differencing[J].Pattern Recognition Letters,2003,24(9):1613-1626.

[7] JOHNSON N F,KATZENBEISSER S.A survey of steganographic techniques[C]//Information hiding.Norwood.Mass:Artech House,2000:43-78.

[8] SANG J.Discrete Fourier transform-based information steganography[J].Journal of Huazhong University of Science & Technology,2008,36(8):5-9.

[9] CHAUMONT M,PUECH W.A DCT-based data-hiding method to embed the color information in a JPEG grey level image[C]//2006 14th European Signal Processing Conference.Florence:IEEE,2006:1-5.

[10] SHEIDAEE A,FARZINVASH L.A novel image steganography method based on DCT and LSB[C]//2017 9th International Conference on Information and Knowledge Technology.Tehran:IEEE,2017:116-123.

[11] DUAN X,SONG H,QIN C.Coverless steganography for digital images based on a generative model[J].Computers,Materials & Continua,2018,55(3):483-493.

[12] 王坤峰,苟超,段艳杰.生成式对抗网络 GAN 的研究进展与展望[J].自动化学报,2017,43(3):321-332.
WANG K F,GOU C,DUAN Y J.Research Progress and Prospect of Generative Adversarial Network GAN[J].Acta Automatica Sinica,2017,43(3):321-332.

[13] YANG J,JIANG Y G,HAUPTMANN A G,et al.Evaluating bag-of-visual-words representations in scene classification[C]// Proceedings of the international workshop on Workshop on multimedia information retrieval.New York:ACM,2007:197-206.

[14] ZHANG Z,LIU J,KE Y,et al.Generative Steganography by Sampling[J].IEEE Access,2019(7):118586-118597.

[15] 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.

[16] 赵丽莉,刘忠喜,孙国强,等.基于非线性状态估计的虚假数据注入攻击代价分析[J].电力系统保护与控制,2019,47(19):38-45.
ZHAO L L,LIU Z X,SUN G Q,et al.Cost analysis of false data injection attacks based on nonlinear state estimation[J].Power System Protection and Control,2019,47(19):38-45.

[17] GULRAJANI I,AHAMED F,ARJOVSKY M,et al.Improved training of wasserstein gans[C]//Advances in neural information processing systems.Vancouver:NIPS,2017:5767-5777.

[18] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al.Generative adversarial nets[C]//Advances in neural information processing systems.Vancouver:NIPS,2014:2672-2680.

[19] ARJOVSKY M,CHINTALA S,BOTTOU L.Wasserstein generative adversarial networks[C]//International conference on machine learning.Sydney:ICML,2017:214-223.

Coverless information hiding based on generative adversarial networks

Duan Xintao,Li Baoxia,Jia Kai,Guo Daidou

(College of Computer and Information Engineering,Henan Normal University,Xinxiang 453007,China)

Abstract:This paper proposes a coverless information hiding method based on generative adversarial networks.The sender uses the camouflage image and the improved wasserstein GAN to generate an image that is visually identical to the secret image. The sender sends the camouflage image,and the receiver uses the camouflage image and the generator to generate an image that is visually identical to the secret image. Compared with the existing information hiding method,this method does not need to modify the carrier. It effectively resists the detection of the steganalysis tool and can achieve the purpose of secure transmission of secret information.

Keywords:Improved Wasserstein GAN;coverless information hiding;information safety

[责任编校　陈留院　赵晓华]