基于双向生成对抗网络的图像感知哈希算法

马 宾[1],王一利[1],徐 健[2],王春鹏[1],李 健[1],周琳娜[3]
(1. 齐鲁工业大学大学(山东省科学院)计算机科学与技术学部,山东济南 250300;
2. 山东财经大学计算机科学与技术学院,山东济南 250014;3. 北京邮电大学网络空间安全学院,北京 100876)

摘 要: 本文提出一种基于双向生成对抗网络(Bidirectional Generative Adversarial Network, BiGAN)的无监督感知哈希生成算法,通过编码网络、生成网络和判别网络间的双向迭代对抗,生成具有较强图像语义特征表示能力的感知哈希码. 本算法通过在编码网络和生成网络间添加跳接层网络结构,将原始图像不同维度的特征信息传递到生成网络,提高生成图像语义学习能力与网络收敛速度;同时,在对抗损失中添加均方误差(Mean Square Error, MSE)损失,增强生成图像的视觉质量与细节表示能力. 最后,基于网络间的多重迭代对抗训练,输出兼备相同来源图像鲁棒性和不同来源图像区分性的高性能图像感知哈希码. 本研究首次采用大型图像数据库进行算法性能评价,实验结果表明,基双向生成对抗网络的感知哈希生成算法与当前其他最新研究方案相比具有更强的版权认证与来源检测能力.

关键词: 感知哈希;生成对抗网络;均方误差;来源检测;哈希码;图像内容认证

基金项目: 国家自然科学基金(No.62272255,No.61872203);国家重点研发计划(No.2021YFC3340600);山东省自然科学基金(No.ZR2019BF017,No.ZR2020MF054);山东省自然科学基金创新发展联合基金(No.ZR202208310038)

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0372-2112(2023)05-1405-08

电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20221224

An Image Perceptual Hash Algorithm Based on Bidirectional Generative Adversarial Network

MA Bin[1], WANG Yi-li[1], XU Jian[2], WANG Chun-peng[1], LI Jian[1], ZHOU Lin-na[3]
(1. *Department of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong* 250300, *China*;
2. *Department of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, Shandong* 250014, *China*;
3. *Department of Cyber Security, Beijing University of Posts and Telecommunications, Beijing* 100876, *China*)

Abstract: An unsupervised perceptual hash generation algorithm based on a bidirectional generative adversarial network (BiGAN) is presented. It generates perceptual hash codes with strong image semantic representation capabilities through bidirectional iterative competition between encoding networks, generation networks, and discrimination networks. Moreover, by adding a skip-connection network structure between the coding network and the generation network, different dimensional features of the original image are transformed from the coding network to the generation network, improving the semantic expression ability of the generated image and convergence speed of the network. At the same time, the mean square error (MSE) loss is added to the network adversarial losses to enhance the visual quality and detail representation ability of the generated image. Finally, a high-performance image perception hash code that possesses the robustness of the same source images and the distinguishability of different source images is obtained via multiple iterative adversarial training networks. A large image database is used for the first time to evaluate the performance of perceptual hash generation schemes in this study. Extensive experimental results show that the proposed algorithm has stronger copyright authentication and source detection capabilities than other state-of-the-art schemes.

Key words: perceptual hash; generative adversarial network; mean square error; source detection; hash code; image content authentication

Foundation Item(s): National Natural Science Foundation of China (No.62272255, No.61872203); National Key R & D Program of China (No.2021YFC3340600); Shandong Natural Science Foundation (No.ZR2019BF017, No.ZR2020MF054);

收稿日期:2022-10-28;修回日期:2023-04-06;责任编辑:覃怀银

Shandong Provincial Natural Science Foundation Innovation and Development Joint Fund (No.ZR202208310038)

1 引言

随着智能终端与数字图像处理技术[1,2]的快速发展,使得图像修改成本不断降低,数字图像的版权验证和来源检测成为图像领域的研究热点. 图像感知哈希作为新兴的多媒体安全技术[3,4],引起了众多学者关注. 图像感知哈希是一种基于图像视觉内容特征生成固定长度散列码的哈希算法[5],其能够将具有相同感知内容的数字图像映射为高度相似的数字摘要,并满足单向性和抗碰撞性. 因而,图像感知哈希算法成为实现图像版权认证和来源检测的理想选择. 一个优秀的图像感知哈希算法应可以实现相同图像感知鲁棒性与不同图像感知脆弱性间的良好平衡.

经典图像感知哈希技术大致可划分为基于统计特征[6~12]、基于局部特征点提取[13~16]、基于频域变换[17~23],以及基于特征降维[24~28]的图像感知哈希算法. 这类算法依赖预先设计的特征提取器生成哈希序列,因而需要广泛的专家知识,而且难以捕捉数字图像内在或抽象的视觉特征,感知哈希算法的性能优化能力不强. 近年来,许多研究开始采用深度学习网络模型生成图像感知哈希码,充分利用不同类型图像的深层特征信息,提升感知哈希算法的鲁棒性和区分性. Li 等人[29]采用分层的方式训练感知哈希网络,降低神经网络的训练的难度,提高算法对内容失真的鲁棒性. Qin 等人[4]通过权重分配策略将两对约束集成到一个整体约束函数中,构造了一种基于多约束卷积神经网络(Convolutional Neural Networks, CNN)的哈希方案.

传统的图像感知哈希方案,通常只能对一种或几种几何变换,例如平移、旋转等具有较好的鲁棒性,而对其他变换,例如仿射变换等,具有很强的脆弱性,导致感知哈希性能不强. 相比传统算法,基于深度学习的感知哈希算法能够取得更好的图像辨识性能,但通常需要大量带标注数据支持模型的训练,且只能对抗较低强度的内容保持攻击. 因而,为了降低深度学习模型对标注数据的依赖性,提升感知哈希算法对抗高强度内容保持攻击的能力,本文提出了一种基于双向生成对抗网络(Bidirectional Generative Adversarial Network, BiGAN)[30]的无监督图像感知哈希生成算法,充分利用 BiGAN 对原始图像隐含特征的强大学习能力,输出能够抵抗多种攻击类型,以及较高攻击强度的图像感知哈希码,从而实现高质量的图像版权认证和来源检测.

2 基于 BiGAN 的图像感知哈希算法

BiGAN 作为一种高效无监督深度模型,可以在不需要了解训练样本底层结构的情况下,对抗学习生成与训练样本一致的数据分布并生成训练样本的隐空间特征表示. 基于 BiGAN 的图像感知哈希生成网络的基本模型由编码网络 E、生成网络 G、联合判别网络 D 和跳接层网络 S 四个子网络组成(如图 1 所示). 其中,编码网络 E 实现从原始图像数据 $\boldsymbol{X}$ 到潜在特征表示 $E(\boldsymbol{x})$ 的映射,其输入为归一化后的训练图像,输出为图像隐空间特征编码. 生成网络 G 将预设的噪声 $\boldsymbol{Z}$ 映射为与目标图像样本一致的数据分布 $G(\boldsymbol{z})$. 联合判别网络 D 区分输入的数据元组是来自编码网络还是生成网络.

针对基础 BiGAN 所存在的生成图像质量不高、输出隐空间特征序列表示能力不足的问题:一方面,通过在编码网络 E 和生成网络 G 之间添加跳接层网络 S,实现编码网络 E 和生成网络 G 之间不同维度的信息传递,将浅层特征与深层特征有机结合,提升生成网络的学习效率和收敛速度,增强感知哈希码的语义信息表示能力,提高生成感知哈希码针对相同来源图像的感知鲁棒性. 另一方面,构造了一种基于均方误差(Mean Square Error, MSE)损失的 BiGAN 性能优化策略,以优化生成图像质量,增强感知哈希码对图像细节的表示能力,提高生成感知哈希码对不同来源图像的区分性. 同时,基于联合判别网络 D 产生对抗损失,反向激励生成网络 G 输出更高质量的生成图像、编码网络 E 输出具有更强图像语义表示能力的隐空间特征编码,从而提高生成图像感知哈希码的质量.

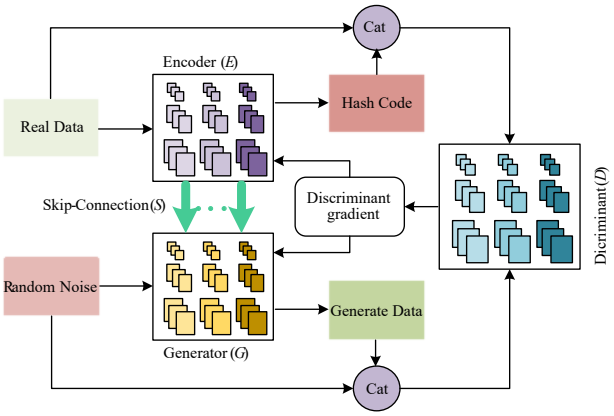

图 1 基于 BiGAN 的感知哈希算法结构图

图 1 中,Real Data 为训练样本图像,Generate Data 代表生成图像.

进一步地,为实现感知哈希码的图像语义表示能力与感知哈希长度之间的优化平衡,实验中采用如式(1)所示的均值量化处理策略[31],在同步增强感知哈希码鲁棒性和区分性的同时,有效降低感知哈希码的长度和复杂度,提高基于感知哈希的图像版权认证与来源

检测效率.

$$H_i(z_i)=\begin{cases}1\ ,\ z_i\geq \mathrm{Mean}(z)\\ 0\ ,\ z_i<\mathrm{Mean}(z)\end{cases} \tag{1}$$

其中,z表示编码网络映射的隐空间特征;z_i表示第i个隐空间特征的值;Mean表示特征值均值计算;$H_i(z_i)$表示对应的第i个隐空间特征量化后的二进制哈希码数值.

2.1　基于BiGAN的图像感知哈希生成网络设计

2.1.1　编码网络

基于BiGAN的感知哈希生成网络中,编码网络E的作用是从原始图像提取隐空间特征信息生成图像感知哈希序列. 本研究中设计编码网络包含10层卷积神经网络,每层卷积神经网络包括图像卷积(Conv2d)、批归一化(BatchNorm)和激活(LeakRelu)三种数据操作处理. 编码网络的初始输入为归一化处理后的训练样本图像. 编码网络E最后一层卷积输出作为图像隐空间特征表示序列. 另外,综合考虑卷积层数量对生成感知哈希码特征表示能力的影响,实验中选择具有10层网络结构的自学习编码网络结构模型,其中包含10个卷积层、9个批归一化层,9个激活层,以保障网络输出具有较强表示能力的隐空间特征编码.

2.1.2　生成网络

生成网络G的目标是将预设的随机噪声分布映射成高质量图像分布,并基于联合判别网络D提供的损失梯度,不断优化生成图像质量. 生成网络G采用全卷积神经网络架构,其中,包含8个反卷积(ConvTranspose2d)层,7个批归一化(BatchNorm)层和7个激活(LeakRelu)层,初始输入为符合高斯分布的随机噪声. 实验中在生成网络G的最后一层使用Tanh激活函数增强生成图像输出细节信息. 生成网络通过逐层反卷积操作,不断增大矩阵维数,最终生成和训练样本图像维数大小一致的图像,同时,通过跳接层S为生成网络G传递样本图像不同维度的特征信息,加快生成网络G的学习能力与收敛速度,增强网络性能.

2.1.3　联合判别网络

联合判别网络D判断数据元组$(\boldsymbol{x},E(\boldsymbol{x}))$和$(G(z),z)$是来自编码网络$E$还是来自生成网络$G$,并输出二者的差异作为误差损失优化生成网络$G$和编码网络$E$联合判别网络$D$的输出,该输出为介于0~1间的数据分布,其对来自编码网络E的数据元组尽量赋高值(靠近1),而对来自生成网络G的数据元组赋低值(靠近0),以优化生成网络G生成与训练样本一致的数据分布,编码网络E输出更具代表性的样本图像隐空间特征表示. 联合判别网络采用梯度上升策略,通过多轮迭代优化提升对输入数据元组的鉴别能力,为生成网络G和编码网络E提供优化的损失梯度.

联合判别网络D采用多层卷积网络提取输入的数据元组特征,并采用Sigmoid激活函数输出数据元组的真实性评价. 判别网络每个卷积层中都包含数据卷积Convolution,激活Active和消除Dropout处理,采用LeakyRelu激活函数. 考虑图像本身数据分布的复杂性,判别网络采用13层卷积神经网络,其中8层用于图像真实性判别,2层用于感知哈希码的真实性判别,3层用于联合编码的真实性判别.

2.1.4　跳接层网络

本研究通过在编码网络E和生成网络G之间添加跳接层网络S,将编码网络E中间层的特征信息链接到生成网络G相应的网络层中(编码网络第5~7层的输出特征链接到生成网络的第2~4层),以传递样本图像深层特征与浅层特征信息,提高基于BiGAN的感知哈希生成网络性能,增强生成图像视觉质量与感知哈希码的特征表示能力,加快网络的收敛速度. 跳接层网络结构将编码网络E与生成网络G的同维度卷积层之间通过跳接层进行链接,将原始图像相同维度的编码特征传递给生成网络G,拼接来自编码网络E的不同卷积层特征信息与生成网络G的对应层特征信息,从而融合不同维度下样本图像的细节与全局信息,生成高质量的目标图像内容分布与更具代表性的特征编码.

跳接层增强网络如图2所示. 其中,卷积运算模块右侧数字表示卷积核个数,上/下方数字表示该层的输入特征图大小. 橙色箭头表示通过跳接层网络将编码网络E所提取的图像特征信息传递并拼接到生成网络G对应维度的卷积层.

2.2　损失函数设计

基于BiGAN的图像感知哈希算法核心是通过训练编码网络E、生成网络G、联合判别网络D以及跳接层网络S,生成具有较强代表性的图像隐空间特征表示,并构造图像感知哈希码,实现相同来源图像感知鲁棒性和不同来源图像区分性. 一方面,联合判别网络D通过梯度上升策略增强对来自器编码网络E和生成网络G数据元组的区分能力,并输出误差损失优化编码网络E和生成网络G. 另一方面,编码网络E和生成网络G采用梯度下降算法不断降低输出元组$(\boldsymbol{x},E(\boldsymbol{x}))$和$(G(z),z)$间的差异,通过BiGAN的多重迭代,最终形成与训练样本高度一致的输出图像分布和具有较强代表性的隐空间特征编码. 为保障BiGAN快速收敛,提高生成图像和感知哈希码质量,本研究设计基于BiGAN的感知哈希生成网络损失函数为:

$$\begin{aligned}\mathrm{Loss}_D=-\mathrm{Mean}[&\log(D(\boldsymbol{x},E(\boldsymbol{x})))\\&+\log(1-D(G(z),z))]\end{aligned} \tag{2}$$

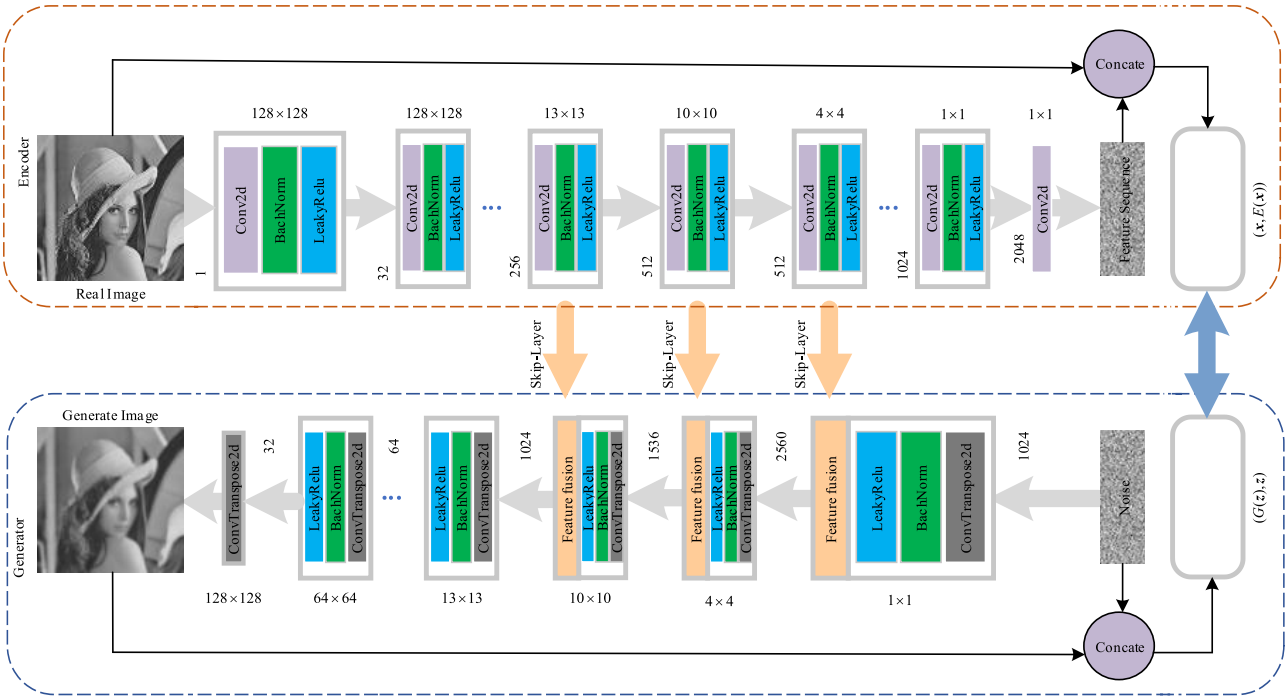

图2 跳接层增强网络结构示意图

$$\begin{aligned}\mathrm{Loss}_{E,G}=&-\mathrm{Mean}[\log(D(G(\boldsymbol{z}),\boldsymbol{z}))\\&+\log(1-D(\boldsymbol{x},E(\boldsymbol{x})))]\\&+\lambda\mathrm{Loss}_{\mathrm{MSE}}\end{aligned}\tag{3}$$

$$\mathrm{Loss}_{\mathrm{MSE}}=\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i-G(\boldsymbol{z}_i))^2\tag{4}$$

式中，$\boldsymbol{x}$是$\boldsymbol{X}$的元素，$\boldsymbol{z}$是$\boldsymbol{Z}$的元素；$\boldsymbol{x}_i$表示第i个训练图像；$\boldsymbol{z}_i$表示第i个随机噪声向量；λ表示本次迭代$\mathrm{Loss}_{\mathrm{MSE}}$的权值系数；Mean表示均值化操作. 实验中采用式(2)优化联合判别网络D，式(2)取值越大，$D(\boldsymbol{x},E(\boldsymbol{x}))$取值越大(靠近1)，$D(G(\boldsymbol{z}),\boldsymbol{z})$取值越小(靠近0)，$\mathrm{Loss}_{\mathrm{MSE}}$越大，也即判别能力越强；采用式(3)优化编码网络$E$与生成网络$G$，式(3)取值越小，$D(G(\boldsymbol{z}),\boldsymbol{z})$取值越大(靠近1)，$D(\boldsymbol{x},E(\boldsymbol{x}))$取值越小(靠近0)，$\mathrm{Loss}_{\mathrm{MSE}}$越小，生成网络$G$输出元组$(G(\boldsymbol{z}),\boldsymbol{z})$与编码网络$E$输出元组$(\boldsymbol{x},E(\boldsymbol{x}))$的相似度越高，也即生成的数据分布具有更多样本图像细节信息，同时，生成的隐空间特征表示包含更多的图像语义特征.

此外，由式(3)可知，本研究在编码网络E和生成网络G的损失函数中增加了MSE损失，增强生成图像与训练样本图像的内容一致性，加速图像的收敛. MSE的值越小，说明预测模型描述实验数据的精确度越高. 因而，在图像生成网络中，采用MSE计算对应像素间的误差，可表征生成图像与原始图像间的一致性程度. 实验中，通过在基于BiGAN的感知哈希生成网络中添加MSE损失，可以为生成网络G提供更好的性能优化梯度，增强生成网络G的细节学习能力，提高生成图像视觉质量，从而对抗激励编码网络输出更具代表性的感知哈希码.

3 实验研究

3.1 实验设计

实验中，首先从图像数据库中随机选择12 000张图像组成训练集，选取8 000张图像作为测试集，以提高基于BiGAN的感知哈希网络所生成感知哈希码的语义表示能力. 同时，考虑到运算效率和计算代价问题，实验中采取了多轮次网络优化训练和性能测试的策略. 具体地，在进行网络训练过程中，每次从训练集中随机选取500张图像进行网络性能优化训练，采用多轮次训练策略(共计24轮次)优化网络性能并保障网络稳定收敛. 在算法性能测试时，每次从测试集中随机选取500张图像进行性能测试(共计16轮次)，以充分验证基于BiGAN的感知哈希码生成算法性能.

另一方面，考虑面部图像特征的复杂性，实验中首先对原始图像进行灰度化并采用双线性插值算法对原始图像进行预处理(将原始图像缩放至128×128的灰度图像)，保留原始图像的关键信息特征. 由于本实验首次采用大规模的CelebAMask-HQ数据库，为准确表达每幅图片的隐空间特征信息，本研究采用1024位感知哈希编码. 实验中采用互相关系数计算不同感知哈希码之间的相关性，并利用受试者操作特征曲线(Re-

ceiver Operating Characteristic curve，ROC)评价基于BiGAN的感知哈希网络图像来源识别性能. 所有实验都选取pytorch框架,并采用配置为Intel 酷睿i9 CPU(72核心、基准速度2.6 GHz),2TB硬盘,英伟达V100 32 GB显存的DeLL R740图形工作站实现.

3.2　基于BiGAN的感知图像哈希算法性能研究

3.2.1　相同来源图像感知鲁棒性验证

为深入探究基于BiGAN的感知哈希生成网络的性能,本实验采用三种不同的网络结构生成图像感知哈希码:基础BiGAN、跳接层增强的BiGAN以及结合MSE损失优化和跳接层增强的BiGAN. 首先,在训练集中随机选取500张图像作为样本. 然后,采用表1所示的攻击策略生成42 500张副本图像,利用上述三种不同的网络结构分别生成图像感知哈希码,计算原始图像与受攻击图像感知哈希的互相关值并求取平均数. 实验结果如表2所示.

表1　不同类型的内容保持攻击

攻击方式	攻击参数	攻击强度
亮度	Photoshop scale	−20,−10,10,20
对比度	Photoshop scale	−20,−10,10,20
高斯低通滤波	标准差	0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0
伽马滤波	gamma	0.7,0.9,1.1,1.3
缩放	比率	0.5,0.6,0.7,0.8,0.9,1.1,1.2,1.3
椒盐噪声	密度	0.001,0.002,…,0.009,0.01
高斯噪声	方差	0.001,0.002,…,0.009,0.01
旋转,裁剪,缩放	旋转角度	1,2,3,4,5
JPEG压缩	质量因子	30,35,…,95,100
水印嵌入	量化步长	0.1,0.2,0.3,0.4,…,0.9,1.0
随机擦除	比率	0.005,0.01,0.015,0.02
中心裁剪	比率	0.05,0.1,0.15,0.2,0.25,0.3,0.35,0.4

实验结果表明,图像在受到不同类型攻击后,基于BiGAN所生成的感知哈希码互相关系数仍然具有较强的相似度. 采用基础BiGAN生成的感知哈希码平均互相关值在0.844 8以上;而当添加跳接层网络结构后,图像感知哈希码平均互相关值进一步改善,达到0.906 1以上,添加跳接层结构明显增强了生成感知哈希码的感知鲁棒性. 即使在使用MSE损失增强图像质量后,受攻击图像生成感知哈希码的鲁棒性受到影响,互相关系数略有下降,但仍然保持在0.862 9以上. 由于图像感知哈希是一个鲁棒性和区分性之间博弈的平衡过程,过高的鲁棒性会导致感知哈希算法区分性能的减弱,反之亦然. 因而,我们在实验中添加跳接层网络的基础上,增加MSE对抗损失以提升图像感知哈希算法的区分能力,实现感知哈希码鲁棒性与区分性的优化调节.

表2　相同来源图像互相关系数

攻击方式	均值		
	BiGAN	Skip-Connection +BiGAN	Skip-Connection +MSE+BiGAN
亮度	0.672 9	0.911 9	0.824 3
对比度	0.973 5	0.949 6	0.931 6
高斯低通滤波	0.989 2	0.985 6	0.963
伽马滤波	0.522 7	0.859 2	0.739 2
缩放	0.985	0.983 8	0.957 8
椒盐噪声	0.994	0.968	0.963 5
高斯噪声	0.983 9	0.914	0.889 2
旋转,裁剪,缩放	0.644 8	0.780 7	0.708 6
Jpeg压缩	0.724 5	0.925 4	0.834 8
水印嵌入	0.998 8	0.998 1	0.997 5
随机擦除	0.902 5	0.907 7	0.889 3
中心裁剪	0.745 8	0.689 2	0.655 5
不同攻击平均互相关	0.844 8	0.906 1	0.862 9

3.2.2　不同来源图像感知鲁棒性验证

为了验证基于BiGAN的感知哈希生成算法识别不同来源图像的能力,首先从测试集中随机抽取了500张图像. 接着,采用表1所示的攻击手段对这些图像进行攻击和篡改. 然后,分别使用三种不同结构的BiGAN生成图像感知哈希码,计算任意两个不同图像感知哈希的互相关值. 每次产生12 500个互相关值. 实验结果见表3.

表3　不同来源图像互相关系数

攻击方式	均值		
	BiGAN	Skip-Connection +BiGAN	Skip-Connection +MSE+BiGAN
亮度	0.254 5	0.119 8	0.082 8
对比度	0.255 1	0.121 3	0.085 8
高斯低通滤波	0.255 2	0.119 6	0.085 8
伽马滤波	0.236 1	0.112 2	0.072 8
缩放	0.221 0	0.122 4	0.083 2
椒盐噪声	0.255 1	0.121 4	0.084 0
高斯噪声	0.255 3	0.123 0	0.083 6
旋转,裁剪,缩放	0.257 9	0.123 2	0.086 3
Jpeg压缩	0.263 0	0.124 2	0.088 1
水印嵌入	0.255 0	0.120 5	0.082 5
随机擦除	0.254 9	0.254 9	0.083 9
中心裁剪	0.245 0	0.133 3	0.084 4
不同攻击平均互相关	0.250 6	0.132 9	0.083 6

结果显示不同图像感知哈希码的互相关值均不大于0.263 0,相比于相同来源的图像,互相关值的取值范围明显下降. 由表3还可以看出,基础BiGAN所产生的感知哈希码互相关系数的均值为0.250 6,基于跳接层增强的BiGAN所产生的感知哈希互相关均值下降到0.132 9,而在增加MSE网络损失后,互相关均值进一步下降为0.083 6. 这是因为添加MSE损失增强了基于BiGAN的感知哈希生成网络对图像细节特征的学习能力,有利于实现感知哈希码在鲁棒性和区分性间的良好平衡.

3.2.3 MSE损失对感知哈希生成网络性能的影响

MSE损失通过计算编码网络E与生成网络G的输出元组之间的差异,一方面,激励生成网络G形成与原始图像高度一致的数据分布,增强生成图像的细节表示能力;另一方面,激励编码网络E输出更具图像语义表示能力的隐空间特征编码. 因而,在基于BiGAN的感知哈希生成网络中,添加MSE损失可以促使网络快速生成更具图像隐特征信息表示能力的感知哈希码;同时,通过增强感知哈希生成网络的学习能力,进一步提升网络收敛速度. 实验中对比了MSE损失对基于BiGAN的感知哈希生成网络收敛速度的影响(实验结果如表4所示). 由实验结果可知,添加MSE损失后,网络的收敛速度明显提升. 当MSE权值系数较小时,对网络收敛速度的优化能力不强,而当MSE权值系数过大时,网络优化梯度下降的步长较大,导致感知哈希生成网络在最优(稳定)工作点附近出现振荡现象,也在一定程度上影响了网络的收敛速度. 综上,当MSE系数位0.5时,基于BiGAN的感知哈希生成网络取得最佳收敛性能,此时所生成的感知哈希码取得最优的图像语义表示能力,感知哈希生成网络具有最高的感知哈希码生成效率.

表4 MSE权值系数对网络收敛速度的影响

MSE权值系数	0	0.1	0.3	0.5	0.8	1.0
收敛轮次	325	245	146	112	123	226

3.3 不同方法间性能比较

相较于经典的感知哈希方案多采取人工设计的算法提取图像的特定特征信息,并基于小样本数据集进行训练,生成图像感知哈希,导致所生成的哈希值只能针对某一种或几种特定类型的攻击具有较强的鲁棒性,而对于其他类型攻击的鲁棒性较差的问题. 本方案首次提出基于大型数据库的无监督深度学习感知哈希生成算法,充分利用BiGAN的深度特征提取能力,生成更具代表性的图像感知哈希,提高输出图像感知哈希的区分性和鲁棒性. 为全面评价本算法的感知哈希性能,实验中分别选取基于矩阵压缩[15]、基于流形学习[27]、以及基于深度学习[29]这几种当前最优秀感知哈希算法与本文所提出的方法进行对比验证,评价本算法针对相同来源图像的感知鲁棒性与不同来源图像的区分性.

在本实验中,我们首先从CelebAMask-HQ测试集中随机选取5 000张图像作为样本. 然后,采用表1所示的攻击策略对原始图像进行内容保持攻击,利用不同的感知哈希生成方法提取图像感知哈希码并计算其互相关性,评估不同感知哈希生成算法实现图像来源检测的性能(实验结果详见表5). 实验结果表明,本研究所提出的算法在图像来源识别方面具有更好的表现,误判率明显低于其他感知哈希生成方法. 当阈值为0.7时,本研究所提出的算法就能够实现不同来源图像的准确检测. 这一结果为相同来源图像的准确识别提供了较大的阈值优化空间,使得基于BiGAN的感知哈希生成算法在图像来源识别方面取得更好的性能.

表5 相同来源图像互相关系数 单位:%

阈值	文献[15]	文献[19]	文献[27]	文献[29]	本方案
0.9	99.98	100	100	100	100
0.8	99.51	100	99.91	99.89	100
0.7	97.15	99.97	99.49	96.73	100
0.6	91.43	99.73	97.99	81.75	99.94
0.5	81.54	98.59	94.06	54.25	99.57
0.4	68.33	94.81	86.12	27.09	98.14
0.3	53.20	85.27	73.10	10.30	93.97

实验中还采用ROC曲线对比不同算法针对相同来源图像的识别能力. 首先对按照表1所示的部分攻击方法(每种方法选取参数最低和最高的两个)产生$5\,000\times12\times2=120\,000$张图像和原始图像混淆处理,采用本文所提出的方法生成感知哈希码并分别计算任意两张图像感知哈希码的互相关值,从0~1调整图像识别分类的阈值,互相关系数大于阈值则认为是相同来源的图像,小于阈值则认为是不同来源图像. 通过实验中真正率(True Positive Rate, TPR)和假正率(False Positive Rate, FPR)的联合分布绘制ROC曲线,对比不同算法间的分类性能,并确定基于BiGAN的感知图像哈希生成算法的最优分类阈值. 实验结果如图3所示,从图中可以看出基于BiGAN的感知哈希生成算法与其他算法相比具有更优秀的分类性能,其ROC曲线取得更靠近坐标图左上角的分布,具有更大的线下面积(Area Under Curve, AUC),也即基于BiGAN的感知哈希算法具有最佳的图像来源识别能力. 实验结果表明:当阈值$\eta=0.79$时,基于BiGAN的感知哈希生成算法取得最佳分类效果. 此时,基于BiGAN生成感知哈希码的图像识别正确率TPR=98.3%,识别错误率FPR=0.76%. 基

于BiGAN的感知哈希生成网络取得针对相同来源图像识别的最好感知鲁棒性,以及针对相同来源图像认证的最好区分性.

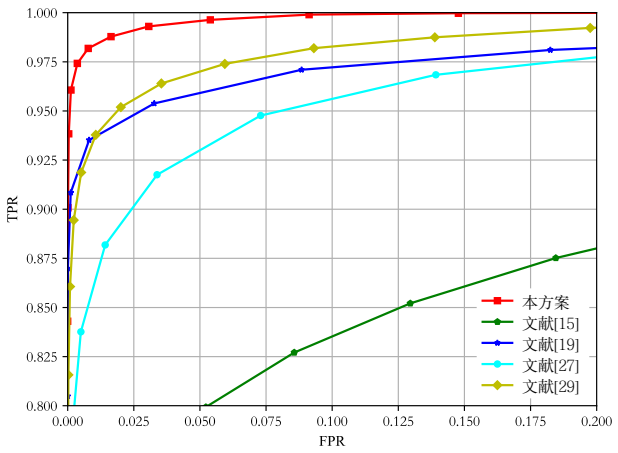

图3　不同感知哈希算法ROC曲线比较

3.4　不同算法运行效率与性能比较

本研究还对比分析了不同算法生成哈希码的效率与性能. 具体方法是:随机选取1 000张图像,分别采用不同算法生成图像感知哈希码,计算生成每个感知哈希码所需的平均时间,并比较其对图像来源正确识别的能力(正确识别不同来源图像和相同来源图像所占总图像数的比值). 实验结果如表6所示,基于BiGAN的感知哈希生成算法所需时间为0.88 s,图像来源正确识别性能为98.94%. 该算法生成感知哈希码所需时间与图像来源识别综合能力明显优于其他大部分方法,实现了图像内容取证能力与算法运行效率之间的最优平衡.

表6　感知哈希生成时间对比表

	文献[15]	文献[19]	文献[27]	文献[29]	本方案
识别性能/%	87.51	96.53	94.12	96.78	98.94
生成时间/s	31.55	7.22	2.56	0.04	0.88

4　结语

本文提出了一种基于双向生成对抗网络的图像感知哈希生成算法实现图像来源检测. 本算法通过在BiGAN中添加跳接层结构链接编码网络和生成网络,提升了图像感知哈希码语义特征表示能力与网络收敛速度. 同时在网络中增加MSE损失,提升图像来源识别与分类的精度. 研究中首次采用大型图像数据库CelebAMask-HQ检验感知哈希生成算法的性能,试验结果表明,与当前最优的感知哈希生成算法相比,本算法可以实现图像来源的准确分类和判定,有效提升图像版权认证和来源检测能力.

参考文献

[1] MA B, SHI Y Q. A reversible data hiding scheme based on code division multiplexing[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(9): 1914-1927.

[2] MA B, CHANG L L, WANG C P, et al. Robust image watermarking using invariant accurate polar harmonic Fourier moments and chaotic mapping[J]. Signal Processing, 2020, 172: 107544.

[3] SRIVASTAVA M, SIDDIQUI J, et al. Local binary pattern based technique for content based image copy detection[C]//2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC). Piscataway: IEEE, 2020: 374-377.

[4] QIN C, LIU E L, FENG G R, et al. Perceptual image hashing for content authentication based on convolutional neural network with multiple constraints[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(11): 4523-4537.

[5] SCHNEIDER M, CHANG S F. A robust content based digital signature for image authentication[C]//Proceedings of 3rd IEEE International Conference on Image Processing. Piscataway: IEEE, 2002, 3: 227-230.

[6] ZHAO Y, YUAN X R. Perceptual image hashing based on color structure and intensity gradient[J]. IEEE Access, 2020, 8: 26041-26053.0-250.

[7] TANG Z J, ZHANG X Q, et al. Robust image hashing with ring partition and invariant vector distance[J]. IEEE transactions on information forensics and security, 2015, 11 (1): 200-214.

[8] SRIVASTAVA M, SIDDIQUI J, ALI M A. Robust image hashing based on statistical features for copy detection[C]// 2016 IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering (UPCON). Piscataway: IEEE, 2017: 490-495.

[9] ZHAO Y, WANG S Z, ZHANG X P, et al. Robust hashing for image authentication using Zernike moments and local features[J]. IEEE transactions on information forensics and security, 2013, 8(1): 55-63.

[10] CHEN Y C, YU W Y, FENG J C. Robust image hashing using invariants of Tchebichef moments[J]. Optik, 2014, 125(19): 5582-5587.

[11] YUAN X R, ZHAO Y. Perceptual image hashing based on three-dimensional global features and image energy [J]. IEEE Access, 2021, 9: 49325-49337.

[12] HOSNY K M, KHEDR Y M, KHEDR W I, et al. Robust image hashing using exact Gaussian-Hermite moments

[J]. IET Image Processing, 2018, 12(12): 2178-2185.
[13] OUYANG J, LIU Y Z, SHU H Z. Robust hashing for image authentication using SIFT feature and quaternion Zernike moments[J]. Multimedia Tools and Applications, 2017, 76(2): 2609-2626.
[14] WANG X F, XUE J R, ZHENG Z Q, et al. Image forensic signature for content authenticity analysis[J]. Journal of Visual Communication and Image Representation, 2012, 23(5): 782-797.
[15] TANG Z J, HUANG L Y, ZHANG X Q, et al. Robust image hashing based on color vector angle and Canny operator[J]. AEU-International Journal of Electronics and Communications, 2016, 70(6): 833-841.
[16] VADLAMUDI L N, VADDELLA R P V, DEVARA V. Robust image hashing using SIFT feature points and DWT approximation coefficients[J]. ICT Express, 2018, 4 (3): 154-159.
[17] LIN C Y, CHANG S F. A robust image authentication method distinguishing JPEG compression from malicious manipulation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(2): 153-168.
[18] TANG Z J, Yang F, HUANG L Y, et al. Robust image hashing with dominant DCT coefficients[J]. Optik, 2014, 125(18): 5102-5107.
[19] HUANG Z Q, LIU S G. Perceptual image hashing with texture and invariant vector distance for copy detection [J]. IEEE Transactions on Multimedia, 2020, 23: 1516-1529.
[20] VENKATESAN R, KOON S M, JAKUBOWSKI M H, et al. Robust image hashing[C]//Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101). Piscataway: IEEE, 2002, 3: 664-666.
[21] TANG Z J, LING M, YAO H, et al. Robust image hashing via random Gabor filtering and DWT[J]. CMC-Computers Materials & Continua, 2018, 55(2): 331-344.
[22] SWAMINATHAN A, MAO Y N, WU M. Robust and secure image hashing[J]. IEEE Transactions on Information Forensics and security, 2006, 1(2): 215-230.
[23] QIN C, CHANG C C, TSOU P L. Robust image hashing using non-uniform sampling in discrete Fourier domain [J]. Digital Signal Processing, 2013, 23(2): 578-585.
[24] KOZAT S S, VENKATESAN R, MIHÇAK M K. Robust perceptual image hashing via matrix invariants[C]//2004 International Conference on Image Processing, 2004. ICIP' 04. Piscataway: IEEE, 2005, 5: 3443-3446.
[25] TANG Z J, ZHANG X Q, ZHANG S C. Robust perceptual image hashing based on ring partition and NMF[J]. IEEE transactions on knowledge and data engineering, 2013, 26(3): 711-724.
[26] TANG Z J, RUAN L L, QIN C, et al. Robust image hashing with embedding vector variance of LLE[J]. Digital Signal Processing, 2015, 43: 17-27.
[27] TANG Z J, LAO H, ZHANG X Q, et al. Robust image hashing via DCT and LLE[J]. Computers & Security, 2016, 62: 133-148.
[28] ZHU X F, LI X L, ZHANG S C, et al. Graph PCA hashing for similarity search[J]. IEEE Transactions on Multimedia, 2017, 19(9): 2033-2044.
[29] LI Y N, WANG D D, TANG L L. Robust and secure image fingerprinting learned by neural network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(2): 362-375.
[30] DONAHUE J, KRÄHENBÜHL P, DARRELL T. Adversarial feature learning[EB/OL]. [2023-04-11]. https://arxiv.org/abs/1605.09782.
[31] HUANG Z Q, TANG Z J, ZHANG X Q, et al. Perceptual image hashing with locality preserving projection for copy detection[J]. IEEE Transactions on Dependable and Secure Computing, 2023, 20(1): 463-477.

作者简介

马 宾 男,1973年出生,山东济宁人. 博士,齐鲁工业大学(山东省科学院)教授. 主要研究方向为多媒体信息安全.
E-mail: sddxmb@126.com

王一利 男,1998年出生,山东济宁人. 齐鲁工业大学(山东省科学院)研究生,主要研究方向为图像感知哈希.
E-mail: a1476529663@163.com

徐 健(通讯作者) 女,1973年出生,山东潍坊人. 硕士,山东财经大学副教授,主要研究方向为多媒体信息安全.
E-mail: sdfixj@126.com