What can Discriminator do? Towards Box-free Ownership Verification of Generative Adversarial Networks

Ziheng Huang[1†], Boheng Li[1†], Yan Cai[1], Run Wang[1*], Shangwei Guo[2],
Liming Fang[3], Jing Chen[1], Lina Wang[1]
[1] Key Laboratory of Aerospace Information Security and Trusted Computing,
Ministry of Education, School of Cyber Science and Engineering, Wuhan University, China
[2] College of Computer Science, Chongqing University, China
[3] College of Computer Science and Technology, Nanjing University of
Aeronautics and Astronautics, China
[†] Equal contribution [*] Corresponding author. E-mail: wangrun@whu.edu.cn

Abstract

*In recent decades, Generative Adversarial Network (GAN) and its variants have achieved unprecedented success in image synthesis. However, well-trained GANs are under the threat of illegal steal or leakage. The prior studies on remote ownership verification assume a black-box setting where the defender can query the suspicious model with specific inputs, which we identify is not enough for generation tasks. To this end, in this paper, we propose a novel IP protection scheme for GANs where ownership verification can be done by checking outputs only, without choosing the inputs (*i.e.*, box-free setting). Specifically, we make use of the unexploited potential of the discriminator to learn a hypersphere that captures the unique distribution learned by the paired generator. Extensive evaluations on two popular GAN tasks and more than 10 GAN architectures demonstrate our proposed scheme to effectively verify the ownership. Our proposed scheme shown to be immune to popular input-based removal attacks and robust against other existing attacks. The source code and models are available at https://github.com/AbstractTeen/gan_ownership_verification.*

1. Introduction

With the rapid development of GANs, we have witnessed fruitful applications of GAN in many fields, such as realistic facial images synthesis [45], fine-grained attribute editing [55], *etc*. Unlike the classification model with specified label prediction, the GANs learn a data distribution and output the synthesized data sample within a certain distribution. In GANs, the discriminator and generator are two essential components, where the discriminator works as a judger to discriminate whether the sample is produced by the generator, the generator learns to generate more realistic samples to confuse the discriminator [16, 7]. Usually, the discriminator is discarded after training since the generator is the core asset for synthesizing high-quality images.

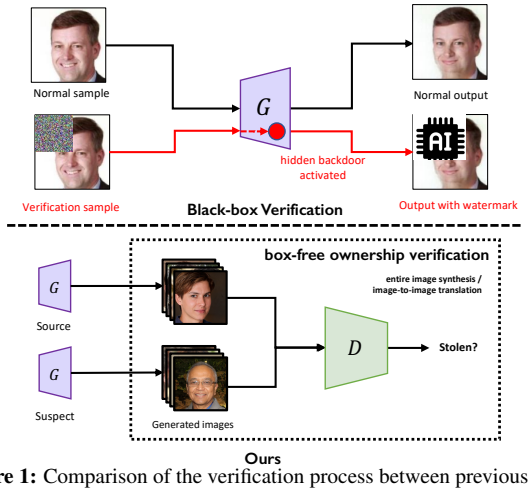

Figure 1: Comparison of the verification process between previous black-box watermark-based verification paradigm [37] and our box-free method. In the black-box setting, carefully-crafted verification samples should be fed to the suspicious model to activate a hidden backdoor in the model (the red circle) and generate watermarked outputs. However, in box-free setting, querying the model with deterministic inputs is not allowed. Ownership verification should be done with only output images.

Training a decent GAN requires a huge investment of resources, such as computing resources, labeled/unlabeled training dataset, time, and human labors [47, 31]. However, well-trained generators are under the threat of unintentional leakage and theft. The adversary may deploy the stolen model on the Internet for profit and the owner (also the *defender*) is only able to verify the ownership remotely by querying the suspicious model [35, 37, 23, 14].

Most existing works on IP protection of DNN models as-

证　书

《慧音 Guardian-物联网语音安全领航者》参赛项目在 2023 年全国高校商业精英挑战赛创新创业竞赛创业计划赛道全国总决赛中，荣获

一等奖

参赛单位名称：武汉大学

优秀辅导教师：王芳　徐晓辉

参赛团队成员：吴定远　蔡艳　范欣　傅彦凯　佘添翼

特发此证，以资鼓励。

中国国际贸易促进委员会商业行业委员会　中国国际商会商业行业商会　中国商业经济学会

二〇二三年十月

中国区国家赛

荣誉证书

武汉大学代表队

在“华为ICT大赛2023-2024”**创新赛**中荣获**三等奖**。

学校名称：武汉大学

选手名称：刘蔚峰、佘添翼、蔡艳

指导老师：汪润

颁发日期

2024年3月31日

HUAWEI ICT Academy

中国战略与Marketing部长
华为技术有限公司

证书编号：ZN202310483

第十六届
中国大学生计算机设计大赛
中南地区赛

参赛队员：刘蔚峰 佘添翼 蔡艳

指导老师：汪润 王丽娜

您创作的《基于深度学习的敦煌壁画修复系统》参赛作品，荣获2023年（第十六届）中国大学生计算机设计大赛中南地区赛**壹等奖**。

中国大学生计算机设计大赛中南地区赛组委会

二零二三年六月

中国大学生计算机设计大赛 中南地区赛 组织委员会

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第11361435号

软件名称：基于深度学习的壁画修复App
[简称：MGN]
V1.0

著作权人：武汉大学

开发完成日期：2023年04月20日

首次发表日期：未发表

权利取得方式：原始取得

权利范围：全部权利

登记号：2023SR0774264

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局
计算机软件著作权
登记专用章

No. 13173531

2023年07月03日

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号： 软著登字第11275761号

软 件 名 称： 基于深度学习的壁画修复系统
[简称：MGN]
V1.0

著 作 权 人： 武汉大学

开发完成日期： 2023年04月18日

首次发表日期： 未发表

权利取得方式： 原始取得

权 利 范 围： 全部权利

登 记 号： 2023SR0688590

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局
计算机软件著作权登记专用章

No. 13061083

2023年06月16日

合作意向书

武汉大学 AISec 团队：

我天津景益天成科技发展有限公司（以下简称我司）意向与你武汉大学 AISec 团队（以下简称 AISec 团队），本着诚信协作、互惠互利的原则，就如下合作事宜达成一致：

一、 合作内容

本协议的合作内容仅限于双方独家合作的“基于深度学习的敦煌壁画修复系统”，以下简称修复系统。

二、 合作方式

我司产品部门协助 AISec 团队，双方共同进行修复系统的落地与推广应用，并为合作团队提供部分技术支持。

三、 权利与义务

1. 为力争使本合作项目早日落地，双方必须充分配合，遵守双方有关约定，履行签约的一切合法商务工作。
2. AISec 团队应充分利用本团队的技术优势，提供推广所需产品，完成算法选代优化，并按本项目的上述规定履行相关责任；我司将充分利用产品部门在软件推广领域的行业优势，协助修复系统的落地，并进行修复系统的区域推广应用。
3. 我司需配合 AISec 团队在产品应用地区建立项目组并完成相关手续，一切费用由双方共同承担。

四、 保密要求

AISec 团队向我司提供有关资料，仅作为本项目签约后的实施时使用，我司承担保密责任；同时 AISec 团队对我司提供的合作资料也承担保密责任。双方均不得将合作资料随意透露给与本项目无关的第三方。

武汉大学 AISec 团队

授权代表签字：刘蒲冬

天津景益天成科技发展有限公司（盖章）

2023 年 8 月 20 日

蓝桥杯大赛

获奖证书

武汉大学蔡艳：

荣获第十四届蓝桥杯全国软件和信息技术专业人才大赛湖北赛区Python程序设计大学A组三等奖。

特发此证，以资鼓励。

证书编号：321400827

证件号码：450502200306021761

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2023年4月23日

国家奖学金荣誉证书

编号：BZK202305187

蔡艳同学荣获 2022 至 2023 学年度本专科生国家奖学金，特颁此证。

中华人民共和国教育部

2023年12月

武漢大學

WUHAN UNIVERSITY

证书编号：XGC202211539

防伪密钥：DE4E8C0324DBA017

防伪中心：https://honor.whu.edu.cn/pc/anti-counterfeit

蔡艳 同学

在 2021-2022 学年度中，学习成绩优异，综合表现突出，被评为**三好学生**。

特发此证，以资鼓励。

武 汉 大 学

二〇二二年十一月

武漢大学

WUHAN UNIVERSITY

证书编号：XGC202205309

查询密钥：4E33DF7A13D90BA9

防伪网站:http://honor.whu.edu.cn/ssp/stu/querycer.aspx

蔡艳 同学

在 2021-2022 年度积极进取，勤奋学习，综合素质全面发展，荣获**甲等奖学金**。

特发此证，以资鼓励。

武 汉 大 学

二〇二二年十一月

490+100+60

武漢大學
WUHAN UNIVERSITY

证书编号：TW202315016

防伪密钥：B1B7DE79D7CE7CBA

防伪中心：https://honor.whu.edu.cn/pc/anti-counterfeit

蔡艳 同学

被评为武汉大学 2022 年度

武汉大学优秀共青团员

特发此证，以资鼓励。

学号：2021302181013

共青团武汉大学委员会

二〇二三年五月

武漢大學
WUHAN UNIVERSITY

证书编号：XGC202313207

查询密钥：6EE8F8532C1266DB

防伪网站:http://honor.whu.edu.cn/ssp/stu/querycer.aspx

蔡艳 同学：

在 2022-2023 学年度中，学习成绩优异，综合表现突出，被评为**三好学生**。

特发此证，以资鼓励。

武 汉 大 学

二〇二三年十一月

武漢大学

WUHAN UNIVERSITY

证书编号：XGC202309147

查询密钥：4B485630048D711D

防伪网站:http://honor.whu.edu.cn/ssp/stu/querycer.aspx

蔡艳 同学：

在 2022-2023 学年度中，学习成绩优异，综合表现突出，荣获**优秀学生甲等奖学金**。

特发此证，以资鼓励。

武 汉 大 学

二〇二三年十一月

编号： CCF-IDEN-CSP-2024-01515

CCF计算机软件能力认证
CERTIFIED SOFTWARE PROFESSIONAL

认证成绩 TRANSCRIPT

基本信息 PROFILE

姓名 /NAME	蔡艳		
注册号/ REG	202414202766	身份证 /ID	450502200306021761
日期 /DATE	2024.03.31	批次 /BATCH	116
编程语言 /PROG.LAN.	C/C++	认证点 /PLACE	武汉大学

认证得分 SCORES

第一题 TASK 1	第二题 TASK 2	第三题 TASK 3	第四题 TASK 4	第五题 TASK 5	总成绩 TOTAL SCORES
100	100	100	0	0	300

排名分析 RANKING

	最高分 HIGHEST	平均分 AVERAGE	满分 FULL	认证人数 TOTAL NO.	排名(前%) TOP %
本次 CURRENT	500	201	500	7672	17.81
累计 OVERALL	500	141	500	222668	3.81

中国计算机学会

2024年04月08日

荣誉证书

蔡　艳　同学

于2022年度积极参与志愿服务活动，表现优异，被评为武汉大学2022年度

优秀青年志愿者

特发此证，以资鼓励！

共青团武汉大学委员会

武汉大学青年志愿者协会

二〇二三年九月