|  |  |
| --- | --- |
| **分** **类** **号：** **TP309** | **单位代码：10183** |
| **研究生学号：2021522052** | **密** **级：公开** |



吉林大学

硕士学位论文

**（学术学位）**

利用数字水印增强隐私计算安全性的研究

Research on Enhancing Privacy Computing Security through Digital Watermarking

**作者姓名：** **赵柏富**

**专** **业：信息与通信工程** **研究方向：信息安全**

**指导教师：温泉** **副教授** **培养单位：通信工程学院**

**2024 年** **5 月**



中国知网 https:Www  cnknet



利用数字水印增强隐私计算安全性的研究

———————————————————————————————

Research on Enhancing Privacy Computing Security through Digital Watermarking

作者姓名：赵柏富

专业名称：信息与通信工程

指导教师：温泉 副教授

学位类别：工学硕士

答辩日期：2024 年 5 月 26 日

中国知网 https:Www  cnknet

利用数字水印增强隐私计算安全性的研究

**摘** **要**

随着大数据时代的到来，数据确权变得十分困难。隐私计算凭借“数据可 用不可见”的特点，受到广泛关注。其中，联邦学习允许数据在本地保留的情 况下参与模型训练，从而保障数据和隐私的安全性。然而，联邦学习也容易受 到保密性攻击，即泄露或窃取用户敏感数据的攻击类型。因此，如何防范这些 攻击，保证数据安全成为一个亟待解决的重要问题。

数字水印具有追踪溯源的功能，利用数字水印能够增强联邦学习的安全性。 在不影响联邦学习的准确性和时效性的前提下，利用数字水印提高联邦学习的 抗攻击能力，具体工作总结如下：

（1）提出一种基于 DWT（离散小波变换）和 SVD（奇异值分解）的双重 鲁棒数字水印算法。首先利用 Arnold 变换对水印图像进行加密，提高了水印图 像的安全性，然后将图像进行离散小波变换，同时使用奇异值分解嵌入双重水 印，增强水印图像的鲁棒性。仿真结果表明，使用该算法后，保证图像的不可 见性的前提下图像的鲁棒性有明显提升。

（2）验证数字水印增强隐私计算安全的可行性。在联邦学习中使用 Non- IID 数据分发和 Adam 梯度下降聚合提升联邦学习的准确率，之后在联邦学习中 应用数字水印方法。仿真结果表明，使用数字水印技术不影响联邦学习的准确 性和时效性，利用数字水印增强联邦学习安全性是可行的。

（3）对利用数字水印增强隐私计算安全性的方法的抗攻击测试。包括 JPEG 压缩攻击、噪声攻击、滤波攻击、缩放攻击、裁剪攻击和旋转攻击。与现 有的几种水印算法相比较，本文提出的算法在多数攻击情景下能够提取出较清 晰的水印图像。当联邦学习遇到保密性攻击导致泄露数据时，即使攻击方对图 片进行处理，但从修改后的图片中仍然可以恢复嵌入的水印，从而实现追踪溯 源的功能，提升联邦学习的安全性。该方法不仅保持联邦学习的准确性和时效 性，还显著提高了其安全性和隐私保护，使联邦学习具有抗攻击能力。

**关键词：**联邦学习，数字水印，小波变换，奇异值分解，Arnold 变换

**Research on privacy calculation methods based on digital watermarking**

**Abstract**

With the advent of the big data era, confirming data ownership has become extremely challenging. Privacy computing, characterized by “usable but invisible data,” has garnered widespread attention. Among these, federated learning allows data to participate in model training while remaining locally stored, thereby safeguarding data and privacy. However, federated learning is also susceptible to confidentiality attacks, which involve the leakage or theft of sensitive user data. Therefore, preventing these attacks and ensuring data security has become a pressing issue.

Digital watermarking, with its traceable functionality, can enhance the security of federated learning. Without affecting the accuracy and timeliness of federated learning, digital watermarking can be used to improve its attack resistance. The key points of the work are summarized as follows:

(1) A dual robust digital watermarking algorithm based on DWT (Discrete Wavelet Transform) and SVD (Singular Value Decomposition) is proposed. First, Arnold transformation is used to encrypt the watermark image, enhancing its security. Next, the image undergoes discrete wavelet transform, and dual watermarking is embedded using singular value decomposition, improving the robustness of the watermark image. Simulation results indicate a significant improvement in the robustness of the image while maintaining its invisibility when using this algorithm.

(2) The feasibility of using digital watermarking to enhance the security of privacy computing is verified. Non-IID data distribution and Adam gradient descent aggregation are employed in federated learning to improve the accuracy of federated learning, followed by the application of digital watermarking methods in federated learning. Simulation results show that using digital watermarking technology does not impact the accuracy and timeliness of federated learning, and enhancing the security

of federated learning using digital watermarking is feasible.

(3) Attack resistance tests on the method of using digital watermarking to enhance the security of privacy computing are conducted. These include JPEG compression attacks, noise attacks, filtering attacks, scaling attacks, cropping attacks, and rotation attacks. Compared to several existing watermarking algorithms, the proposed algorithm can extract relatively clear watermark images in most attack scenarios. When federated learning faces confidentiality attacks leading to data leakage, even if the attacker processes the image, the embedded watermark can still be recovered from the modified image, thereby achieving traceability and enhancing the security of federated learning. This method not only maintains the accuracy and timeliness of federated learning but also significantly improves its security and privacy protection, endowing federated learning with attack resistance capabilities.

**Key words:** Federated learning, Singular value decomposition, Arnold transform， Digital watermarking, Wavelet transform

目 录

[第 1 章 绪论 1](#bookmark1)

[1.1 课题研究背景及意义 1](#bookmark2)

[1.2 国内外研究现状 2](#bookmark3)

[1.2.1 联邦学习 2](#bookmark4)

[1.2.2 数字水印 3](#bookmark5)

[1.3 论文的研究内容及章节安排 4](#bookmark6)

[第 2 章 隐私计算和数字水印理论研究 7](#bookmark7)

[2.1 隐私计算的基本原理 7](#bookmark8)

[2.1.1 隐私计算 7](#bookmark9)

[2.1.2 联邦学习 8](#bookmark10)

[2.2 数字水印的基本原理 11](#bookmark11)

[2.2.1 数字水印的基本框架 12](#bookmark12)

[2.2.2 数字水印的分类 12](#bookmark13)

[2.2.3 离散小波变换 14](#bookmark14)

[2.2.4 数字水印的评价指标 16](#bookmark15)

[2.3 本章小结 18](#bookmark16)

[第 3 章 基于 DWT-SVD 的双重鲁棒数字水印算法 19](#bookmark17)

[3.1 基于奇异值分解(SVD)的鲁棒数字水印算法 19](#bookmark18)

[3.1.1 鲁棒数字水印算法 19](#bookmark19)

[3.1.2 奇异值分解 20](#bookmark20)

[3.1.3 基于奇异值分解的鲁棒数字水印算法 20](#bookmark21)

[3.2 基于 DWT-SVD 的鲁棒水印算法 23](#bookmark22)

[3.2.1 单奇异值 S 水印算法 23](#bookmark23)

[3.2.2 单奇异值 U 水印算法 25](#bookmark24)

[3.2.3 单奇异值 V 水印算法 26](#bookmark25)

[3.3 基于 DWT-SVD 的双重奇异值水印 27](#bookmark26)

[3.3.1 双重奇异值 S 、U 水印算法 28](#bookmark27)

[3.3.2 双重奇异值 S 、V 水印算法 30](#bookmark28)

[3.3.3 双重奇异值 U 、V 水印算法 31](#bookmark29)

[3.4 水印预处理 33](#bookmark30)

[3.4.1 Arnold 变换 33](#bookmark31)

[3.4.2 水印预处理实验仿真 34](#bookmark32)

[3.5 实验仿真平台搭建 36](#bookmark33)

[3.6 基于 DWT-SVD 的双重鲁棒数字水印的算法实现 36](#bookmark34)

中国知网 https:Www  cnknet

[3.6.1 水印的嵌入 36](#bookmark35)

[3.6.2 水印的提取 38](#bookmark36)

[3.6.3 实验仿真 39](#bookmark37)

[3.7 本章小结 39](#bookmark38)

[第 4 章 利用数字水印增强隐私计算安全性的可行性研究 41](#bookmark39)

[4.1 横向联邦学习模型 41](#bookmark40)

[4.1.1 Mnist 数据集 43](#bookmark41)

[4.1.2 数据分发问题 43](#bookmark42)

[4.1.3 实验设计 45](#bookmark43)

[4.1.4 Adam 梯度下降 47](#bookmark44)

[4.2 实验仿真平台搭建 47](#bookmark45)

[4.3 利用数字水印增强隐私计算的功能性测试 48](#bookmark46)

[4.4 本章小结 49](#bookmark47)

[第 5 章 利用数字水印增强隐私计算的抗攻击测试 51](#bookmark48)

[5.1 实验仿真平台搭建 51](#bookmark49)

[5.2 抗攻击测试结果与分析 51](#bookmark50)

[5.2.1 JPEG 压缩攻击 51](#bookmark51)

[5.2.2 噪声攻击 53](#bookmark52)

[5.2.3 滤波攻击 55](#bookmark53)

[5.2.4 缩放攻击 57](#bookmark54)

[5.2.5 裁剪攻击 58](#bookmark55)

[5.2.6 旋转攻击 59](#bookmark56)

[5.3 抗攻击性能对比分析 61](#bookmark57)

[5.4 本章小结 62](#bookmark58)

[第 6 章 总结与展望 63](#bookmark59)

[6.1 全文总结 63](#bookmark60)

[6.2 未来展望 63](#bookmark61)

[参考文献 65](#bookmark62)

中国知网 https:Www  cnknet

**第** **1 章** **绪论**

**1.1 课题研究背景及意义**

随着算法的不断创新和训练数据的增加，机器学习技术特别是深度学习 技术在人工智能应用领域取得了巨大成功[1]。例如，在图像识别领域，卷积 神经网络已经超越了人类的识别准确率；在自然语言处理领域， BERT 算法 刷新了多项自然语言处理纪录[2] ；在推荐领域， 科技公司如 YouTube 、 Facebook 和 Netflix 利用智能推荐引擎提升了用户黏性和留存率[3-5]。

然而，当前深度学习的成功离不开大规模的训练数据[6] 。随着移动互联 网的普及，数据规模不断扩大且变得越来越复杂。数据不仅是训练模型所需 的输入，更是为企业带来经济收益的无形资产。虽然构建大型公共项目数据 集的机构越来越多，但由于成本等因素，大型开源数据库的构建仍然困难。

由于数据不仅包含数值属性还具有财产属性，各国和企业越来越重视数 据安全并将其保护起来，这使得数据共享变得愈发困难。虽然越来越多的机 构致力于构建大型公共项目数据集，例如用于视觉对象识别软件研究的大型 图像数据库项目 ImageNet[7] ，但是基于现实的考量使得大型开源数据库的构 建困难重重。

近年来，涉及个人和公众的隐私泄露事件愈演愈烈。例如，2018 年发生 在 Facebook 上的数据泄露事件引发了公众对数据隐私的担忧。人们对于隐私 保护和信息安全的期望越来越高，各个国家和地区纷纷出台了保护政策，致 力于解决不良事件，构造安全的网络生态。2022 年 1 月，国务院发布《“十 四五”数字经济发展规划》，提出要加强高质量数据要素供给，发挥数据要素 作用，推动数据的流通与开发利用。然而， 在保护数据安全方面仍然困难重 重，因此，隐私计算成为破解数据供给和应用难题的关键。

在面对上述问题时，隐私计算中的联邦学习技术显得尤为重要。因此， 为了应对“数据孤岛”的问题，McMahan 等人[8]在 2016 年引入了联邦学习 的概念，联邦学习是一个基于分布式数据集的机器学习框架。 各终端以中央

1

中国知网 https:Www  cnknet

服务器为核心构建联邦学习框架，共享模型的相关信息，但本地训练数据只 保留，不会在终端之间共享。联邦学习既可以满足用户对隐私安全的担忧， 又可以符合相关法规，在最大程度上挖掘数据、机器建模。

尽管联邦学习具有巨大的商业价值，但也面临着许多问题和挑战。近年 来，针对联邦学习的攻击方式有很大发展，主要包括模型反转攻击[9]、投毒 攻击[10-11]、对抗样本攻击[12] 等，按照攻击的重点不同又可分为保密性攻击、 完整性攻击和可用性攻击[13]。

完整性攻击旨在利用毒化模型大幅影响联邦学习的准确度，使模型产生 错误的训练输出。典型的完整性攻击包括标签翻转[14]和后门攻击[15] ，它们通 过训练误导目标模型，使其输入攻击者指定的预测结果[16]。

可用性攻击旨在攻击分类的可用性。通过恶意后门、对抗样本等方法[17]， 使联邦学习中的目标模型无法使用。

保密性攻击旨在窃取用户敏感数据。这类攻击试图窃取本地训练数据或 反向推断出训练模型[18] ，但攻击者不会干扰训练进度或修改模型，而是通过 扮演参与方的角色间接发起攻击。

针对上述攻击方法，为了提高联邦学习的保密性和安全性，本文提出了 一种利用数字水印增强隐私计算安全性的方法。数字水印技术凭借其可证明 性、不可感知性和鲁棒性，可以防御和追踪溯源，很好地应对联邦学习的保 密性攻击。

**1.2 国内外研究现状**

**1.2.1 联邦学习**

目前，隐私计算在理论上愈发成熟， 许多将联邦学习和隐私保护相融合衍 生的技术具有更加广阔的应用场景。联邦学习与可信传输环境以区块链为代表 的结合，受到国内外学者的极大关注[19-22]。其中，Kim 等人[23]提出了基于区块链 的联邦学习，用以实现无需集中的训练数据和协调、缩小性能差距的目标；Qu 等人[24]也提出了基于区块链的联邦学习，除了缩小性能差距，还利用雾计算实 现分散隐私；Pokhrel 等人[25]提出了基于自主区块链的联邦学习，使用分布式 交换与验证对本地车载机器学习进行模型更新，实现了隐私感知并提高了车载

2

中国知网 https:Www  cnknet

通信网络的可靠性；赵军等人[26]设计了一种基于隐私保护的区块链联邦学习系 统，通过信誉机制帮助家电制造商根据客户数据训练机器学习模型。

训练模型是联邦学习安全中重要的组成部分。因此，IEEE 联邦学习协会在 2019 年组织开展了关于数据隐私和保密的研讨会议，规划了联邦学习的安全测 评与评级，并展示了符合标准的文章。其中，Li 等人[27]提出的联合学习中异常 客户端行为检测，填补了联邦学习安全检测领域的空白；Ramanan 等人[28]提出 了一种基于区块链网络的联邦学习架构，使用区块链技术辅助模型并增强联邦 学习的安全；Lu 等人[29]在 2019 年设计了一个基于区块链的安全数据共享架构， 通过区块链技术对联邦学习模型的数据隐私保护进行辅助；2020 年，Lu 等人[30] 在之前的基础上提出了一种基于联邦学习的新方案，采用异步联邦学习来减轻 传输负载并解决提供商的隐私问题，利用深度强化学习 (DRL)进行节点选择为模 型传输提高效率。

近年来，为了抵御联邦学习中的保密性攻击，在联邦学习中使用数字水印 技术得到了广泛应用。例如，Tekgul 等人[31]提出了 WAFFLE，使用联邦学习训练 水印 DNN 模型，在每次将本地模型聚合到全局模型之后，引入了一个重新训练 步骤。Han[32]提出了一种基于联邦学习的鲁棒零水印方案，通过联邦学习训练稀 疏自编码器网络，应用经过训练的稀疏自编码器网络从皮肤病医学图像中提取 图像特征，并对图像特征进行二维离散余弦变换，选择低频变换系数以创建零 水印，以解决远程皮肤病学医疗保健框架的隐私和安全问题。 Yang[33]提出了一 种新的水印方案来解决安全联邦学习中的版权保护问题，通过从客户端嵌入一 个后门到联邦学习模型中，为发起者提供了一种联邦学习模型版权保护方法。

综上所述，在联邦学习结合数字水印技术方向，现有方法主要集中在模型 水印和零水印两个方面，通过这些方法可以提升联邦学习的安全性。然而，保 密性攻击主要针对本地数据，因此在联邦学习训练开始前对本地数据进行加密 操作可以很大程度上抵御这些攻击。数字水印的追踪溯源和抗攻击能力有助于 联邦学习应对保密性攻击，因此本文计划利用数字水印增强联邦学习的安全性。

**1.2.2 数字水印**

数字水印技术自上个世纪末开创以来，已成为当前的研究热点问题之一。

3

中国知网 https:Www  cnknet

学者们提出了大量的水印算法，可用于图像、文本、音频和视频等数字产品的 版权保护或内容认证。最早的一篇数字水印论文是 Tirkel 等人[34] 发表的 “Elecrtronic Watermark”，在该论文中，作者提出了两种在数字图像中嵌入 水印的方法，都是通过修改 LSB 位平面嵌入水印，以达到增加安全性的目的。

数字水印算法根据原始载体是否做过变换分为时/空域算法、压缩域算法和 变换域算法三类。Mansouri 等人[35]提出了一种基于内容密钥的压缩域安全视频 水印算法。Sun[36]提出了一种基于压缩域的可伸缩视频水印算法，将原始水印重 构到不同的尺度上，调整水印的系数，在离散余弦子块中嵌入重构后的水印， 并设计了密钥控制机制， 自适应地选择子块。在变换域算法研究方面，Singh[37] 提出了利用 CKGSA(Chaotic Kbest Gravitational Search Algorithm)的 DCT- SVD 域优化鲁棒水印算法。Li 等人[38]提出了基于四元数小波变换和张量分解的 彩色遥感图像数字水印方案，该方案利用四元数小波变换(QWT)和张量分解，构 造了一种同时考虑频率和统计模式特征的彩色遥感图像鲁棒非盲水印方案。 Chen 等人[39]提出了结合加密技术和数字水印技术的医学图像共享方法，水印图 像由权威诊断结果与原始医学图像的哈希值串联生成的二维码图像作为水印图 像，用离散余弦变换算法嵌入水印图像，脱敏后的水印医学图像被存储到区块 链中。Sathya 等人[40]采用非冗余帧识别 NRFI 算法将冗余帧和非冗余帧分离到 特定的数据库，采用关键帧选择算法来识别合适的关键帧，利用离散小波变换 将关键帧分解为子带，将水印图像的主分量嵌入到视频中已识别的关键帧中。

目前，基于变换域的数字水印算法研究最为广泛，而对于同时嵌入两重水 印的版权保护和内容认证方法研究较少。本文将对双重水印算法进行研究，构 造一种双重鲁棒数字水印算法，用于对数字图像进行版权保护。

**1.3 论文的研究内容及章节安排**

联邦学习容易受到各种攻击，其中保密性攻击尤为突出。保密性攻击是针 对隐私数据的一种攻击方式，该类攻击不仅试图泄露、窃取本地数据，还试图 暴露隐私数据并反向推断出训练模型，对联邦学习危害较大。数字水印凭借其 可证明性和鲁棒性，能够提高隐私计算自身的安全性，充分应对在联邦学习中 遭遇的保密性攻击。该方法需要在不影响联邦学习的准确性和时效性的前提下， 4

中国知网 https:Www  cnknet

利用数字水印提高联邦学习的安全性和保密性同时使联邦学习具有抗攻击能力。 因此提出一种基于 DWT-SVD 的双重鲁棒数字水印算法并应用到联邦学习中，

通过验证其可行性和抗攻击能力。

全文共有六个章节，各章节安排如下图所示：

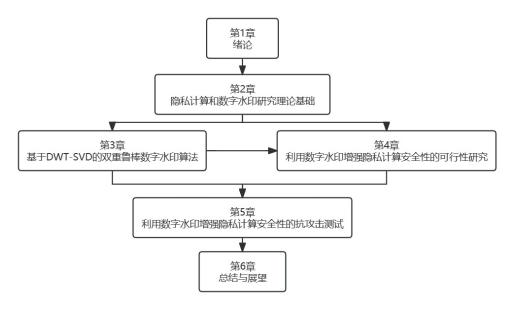


图 1.1 论文章节安排

第 1 章，绪论。绪论首先介绍了联邦学习和数字水印的研究背景和重要性， 包括联邦学习受到的攻击类型，并探讨了将数字水印融入联邦学习以抵御攻击 的方法。通过分析数字水印，提出了一种基于 DWT-SVD 的双重鲁棒数字水印 算法。最后，介绍了本课题的研究内容和论文的结构安排。

第 2 章，隐私计算和数字水印研究理论基础。本章介绍了隐私计算和数字 水印的基本概念，为解决联邦学习的保密性攻击提供了理论支持。接着介绍了 四种主要的隐私计算方法，并着重介绍了联邦学习实现方式。 并通过介绍 PSNR 、SSIM 和 NC 三种评价数字水印的计算公式，为验证本文提出数字水印 算法的有效性搭建了基础。

第 3 章，基于 DWT-SVD 的双重鲁棒数字水印算法。本章首先通过离散小 波变换对图像进行处理，以提高水印图像的隐蔽性。接着，使用 Arnold 变换对 水印图像进行加密，增强了水印的安全性。通过奇异值分解技术，进一步提升 了水印的鲁棒性。仿真实验表明，该算法在不可见性和鲁棒性方面相较于基于 SVD 的水印算法和基于 DWT-SVD 的鲁棒水印算法有显著改进。

5

中国知网 https:Www  cnknet

第 4 章，利用数字水印增强隐私计算安全性的可行性研究。本章通过调整 非独立同分布（Non-IID）的数据分发方式和使用 Adam 梯度下降聚合方法来提 高联邦学习的准确率。在调整后的联邦学习模型中，应用第 3 章提出的数字水 印技术，实验结果显示，该技术不仅保持了联邦学习的准确性和效率，还增强 了其安全性，证明了利用数字水印提升联邦学习安全性的可行性。

第 5 章，利用数字水印增强隐私计算安全性的抗攻击测试。在第 4 章的基 础上对联邦学习算法进行了抗攻击测试，包括 JPEG 压缩攻击、噪声攻击、滤 波攻击、缩放攻击、裁剪攻击和旋转攻击。实验结果表明，在大多数攻击情况 下，提取出的水印图像效果良好。因此，该方法不仅保持了联邦学习的准确性 和效率，还提高了其安全性和保密性，并赋予了抗攻击能力。

第 6 章，总结与展望。本章总结了前 5 章所做的工作，并总结现有的不足， 对未来的工作进行了展望。

6

中国知网 https:Www  cnknet

**第** **2 章** **隐私计算和数字水印理论研究**

**2.1** **隐私计算的基本原理**

**2.1.1 隐私计算**

隐私计算的核心技术涵盖软件层面的多方安全计算、同态加密[41]和联邦学 习[42] ; 在硬件层面则主要涉及可信执行环境。

(1)安全多方计算(Secure Multi-Party Computation，SMC) 是一种由众多 参与方共同参与计算的隐私计算方法，具有保护数据安全、计算平等性和去中 心化的特点。安全多方计算打破了基于共享数据的传统计算方式，以协议为计 算主体代替所有参与者进行安全的多方计算。在进行多方计算的全过程中，各 参与方只能获得私有数据的输出结果，而对其他参与方的信息一无所知。同时， 协议确保了所有数据分享者权利和地位的平等性，任何数据所有者都有权启动 计算任务并获得基于全部数据的精确运算结果。安全多方计算主要的分支方向 有不经意传输、混淆电路和秘密分享，在现实应用场景下，可以自由组合实现 隐私保护的目的。

(2)同态加密(Homomorphic Encryption，HE) 允许用户对已经加密的信息 进行处理和运算。加密数据可以执行任何明文操作，而不必解密或加密数据， 凭借其特性使得用户对加密信息进行深入且不受限的处理和分析，同时可以大 幅度减少数据加密解密的次数，在实际应用过程中既能减少处理代价，又能节 省通信开销。同态加密具有加法同态性和乘法同态性，一些算法仅能具备其中 一种性质，而全同态加密可支持对密文进行有限次数的同态加法和同态乘法运 算[43]。

(3)联邦学习(Federated Learning，FL) 是一种新的深度学习算法，它将 参与者的本地学习数据集成到中央服务器中，服务器将数据处理和学习结合起 来，然后创建联邦学习模型并将其发送给参与者，多次迭代直至模型收敛。联 邦学习全程把终端数据存储在本地，仅通过传输模型进行训练，通过“数据不 出厂”的方式最大程度地避免数据泄露。在应用过程中，依据用户与特征信息

7

中国知网 https:Www  cnknet

的重合度，联邦学习可被划分为横向、纵向和迁移学习［44］等不同类型。

(4)可信执行环境(Trusted Execution Environment，TEE) 是一种基于可 信硬件的隐私保护技术。将数据分为普通数据与敏感数据两类，并针对两种数 据分别划分存储和运算环境使之隔离，在软件层面设置加密和访问权限，在硬 件层面划分单独的存储区域，所有关键数据的计算都在这个隔离的内存空间中 进行［44］。当面临攻击时，可信执行环境可以在机密数据存储和软件执行算法两 个方面保护其不受干扰和窃取，从而全面保护隐私数据的完整性和可靠性。

由于技术路径的差异，上述四种隐私计算技术都具有各自的技术特点和优 缺点，适用于不同的应用场景，如表 2.1 所示。

表 2.1 隐私计算技术的特点

|  |  |  |  |
| --- | --- | --- | --- |
| 名称 | 技术特点 | 优势 | 劣势 |
| 安全多方计 算 | 不暴露隐私， 进行联合 | 计算隐私性和适 用性好 | 通信开销大，效率有待提 升，扩展型不足 |
| 同态加密 | 对密文进行计 算 | 隐私性高 | 计算、存储开销大，效率较 低 |
| 可信执行环 境 | 引入可信硬件 | 易开发、保密性 能佳 | 通用性和可靠性低，依赖硬 件 |
| 联邦学习 | 多方联合训 练，优化模型 | 计算精度高 | 通用性和效率低，保密性一 般 |

安全多方计算和同态加密在保护隐私上表现优异，但二者计算复杂度高、 通信开销大、对性能要求高并且扩展性不足，实际应用过程中易受到多方面掣 肘，因此难以支持大规模应用；基于可信执行环境的方案，凭其物理层面隔离 的特性，在隐私保护上表现出高可靠性和安全性，但硬件设计和制造成本高， 局限性较大；联邦学习的通信开销较小，易大规模开展，但隐私保护方面稍显 不足，用户隐私仍可能受到攻击。综上所述，每种技术都有其优势和局限性， 在实际应用中应结合各自的应用场景，选择合适的优化方法，发挥隐私保护性 能的最大潜力。

**2.1.2 联邦学习**

为了更准确地阐述联邦学习框架的实施细节，以下是该框架的详细流程：

假定K个客户端{Qi 1 想要完成对于数据的分析，他们都有各自的隐私数据 {pi 1。使用传统的机器学习，客户端需要把隐私数据{pi 1 上传到中央服务

器中，才可以进行机器学习的过程。但是在隐私数据上传的过程中，很可能发 生隐私数据的泄露，无论在传输过程还是保存在中央服务器的数据都并不是安 全的做法，数据离开本地之后，随时都会面临泄露的风险，这就导致了传统的 机器学习的工作方式不能给使用者提供安全保护。为了保护用户的隐私安全，

联邦学习可以让用户的隐私数据只在本地进行训练，无需把隐私数据上传到中 央服务器的过程。那么联邦学习的具体工作原理是：联邦学习的中央服务器会 把全局模型MFL 分发给所有需要进行训练的客户端{Qi 1 ，在接收到全局模型 后，客户端将自己的隐私数据在本地使用全局模型训练，可以得到各自的本地 局部训练模型{Mi 1 ，与传统的机器学习不同的是， 客户端只需将得到的 {Mi 1 上传到中央服务器即可，中央服务器在接收到所有客户端上传的{Mi 1， 会对MFL 进行聚合更新，从而得到新的全局模型MFL。并不断循环这个过程，这 样会使MFL 更新的越来越精确

假设传统的机器学习模型MSUM 的准确率VSUM 、联邦学习模型MFL 的准确率 VFL ，可以得到:

∣ VFL − VSUM ∣< δ ....................（2.1）

δ是一个很小的数，这个公式的意义是可以将传统机器学习的准确率和联邦 学习的准确率进行逼近，如果δ足够小，可以说明对于传统的机器学习和立邦学 习来说，两者的功能是相近的，联邦学习不出本地的训练方式足以体现出其优 越性。

联邦学习的流程：中心服务器把模型参数wt 发送给所有需要联邦学习进行 训练数据的客户端{Qi 1 ，在接收到wt 后，客户端进行本地的数据训练，同时 可以得到训练的局部模型：

 ∑u∈P i f  ................. 

在此过程中，f表示损失函数， pi 代表第i个客户端的隐私数据。客户端根 据得到的局部模型Fi (wt )可以得到局部模型的梯度：

 ∑u∈Pi ∇f ................ 

所有客户端把局部模型的梯度gi (wt )都上传到中央服务器上，中央服务器 收到的所有局部模型的梯度进行汇总，之后实现对联邦学习模型MFL 的优化：

9

中国知网 https:Www  cnknet

w ∗ = argminF(w) .................... (2.4)

 ................... 

wt+1 = wt − ∑i∈k gi  ................ (2.6)

wt+1代表优化后的联邦学习模型，w ∗ 是对应的参数，η代表学习率。

上述提到的K个客户端{Qi 1 都有各自的隐私数据{pi 1 ，其中pi 是隐私数 据集合，Pi 由三部分(I, X, Y)构成，一部分是数据记为I ，另一部分是特征记为X， 最后一部分是标签记为Y 。数据I 、特征X和标签Y构成了一个客户端的数据集记 为pi = (I, X, Y) 。在联邦学习中，根据不同的数据拥有者的数据特征空间和样本

ID 空间的重叠关系的不同，可以将联邦学习分为以下三种类型[44]：横向联邦学 习（ Horizontal Federated Learning ，HFL）， 纵向联邦学习 （ Federated Transfer Learning，FTL），联邦迁移学习（Vertical Federated Learning， VFL）。

（1）横向联邦学习[45] ，适用于数据集在数据特征上有重叠，但数据样本不 同的联邦学习参与方，可以表述为：

Xi ≠ Xj, Yi ≠ Yj, Ii = Ij, ∀pi, pj, i ≠ j.................................... (2.7)

横向联邦学习框架图如下图所示：

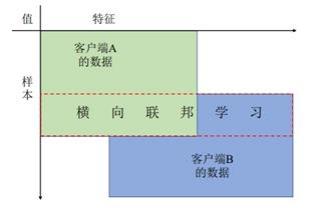


图 2.1 横向联邦学习

（2）纵向联邦学习[46]，适适用于训练数据在数据样本上有重叠，但数据特 征不同的联邦学习参与方，分可以表述为：

Xi = Xj, Yi = Yj, Ii ≠ Ij, ∀pi, pj, i ≠ j .............. (2.8)

10

中国知网 https:Www  cnknet

纵向联邦学习框架图如下图所示：

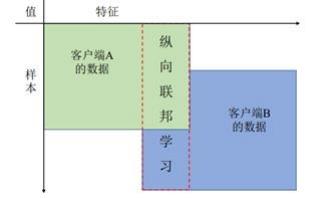


图 2.2 纵向联邦学习

（3）联邦迁移学习[47] ，适用于参与方的数据样本和数据特征都很少重叠的 情况，可以表述为：

xi ≠ xj, yi ≠ yj, Ii ≠ Ij, ∀pi, pj, i ≠ j .............. (2.9)

联邦迁移学习框架图如下图所示：

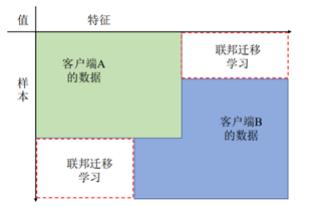


图 2.3 联邦迁移学习

**2.2** **数字水印的基本原理**

数字水印的嵌入过程包括水印的预处理、载体图像变换以及将水印信息嵌 入到载体图像三个阶段。预处理阶段主要对水印进行加密处理，降低水印的相 关性。基于变换域的载体图像处理需要对载体图像进行变换域操作，使载体图 像能够嵌入水印。嵌入水印信息的方法有很多种，比如基于图像像素的操作， 通过修改像素值使其与原始图像产生差异来嵌入水印信息。

对水印的提取过程通常采用嵌入过程的逆变换法，对于一幅嵌入水印的图

11

中国知网 https:Www  cnknet

像经过一定程度上的改变后，仍能从图像中提取出水印信息，即水印能够抵御 对嵌入后数据的一定操作，不因一些细微的操作磨灭。可以数字水印评价指标 包括不可见性、鲁棒性等来评价性能优劣。

**2.2.1 数字水印的基本框架**

数字水印基本框架通常由两部分组成[48] ：水印嵌入系统和水印提取框架。 如下图所示：

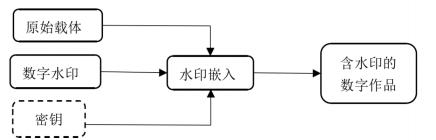


图 2.4 数字水印嵌入框架

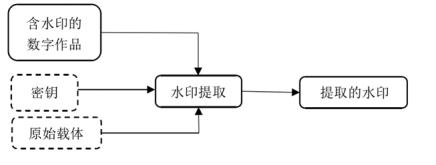


图 2.5 数字水印提取框架

**2.2.2 数字水印的分类**

到目前为止，已经有众多类型的数字水印加密算法，根据水印的可见性对 水印进行区分，即可见水印和不可见水印。

可见水印（Visible watermark）是指人眼可以看见的水印，类似于人为加 在图像上的标识，其所包含的信息可在观看图片或视频时同时被看见，可见水 印在图像中应用广泛，也可用于音频和视频中，用在音频中则称为可听水印。

不可见水印（ Invisible watermark）：是一种更广泛应用的水印，隐蔽性 强，难以察觉，期望能借此避免或阻止数字媒体未经授权的复制和拷贝，但在 版权纠纷时，版权所有者可以通过提取标记来证明所有权。不可见水印又可分 为鲁棒性水印和脆弱性水印。

12

中国知网 https:Www  cnknet

（1）脆弱数字水印

脆弱数字水印是一种当遭受攻击之后，通过对水印可以进行检测是否遭受 过攻击的算法。脆弱数字水印根据嵌入水印的位置不同可分为作用于空间域的 和作用于变换域的。

空间域的脆弱数字水印算法大多是修改图像的像素最低有效位(LSB) [49]。 例如，S．Walton 提出的 Checksum 算法，基于图像像素值高 7 位的校验和。

变换域的脆弱数字水印算法是通过对变换域系数修改而产生的。 这类方法 一般是对图像进行局部或整体的变换，常用的变换有离散余弦变换[50]、离散小 波变换[51]等。很多学者都参与了基于变换域的脆弱水印算法研究：Marvel 等人 提出了一种基于分块的 DCT 算法，利用伪随机发生器选择嵌入水印的位置，利 用 DCT 块和密钥构造水印。

空间域的脆弱数字水印算法和变换域的脆弱数字水印算法在本质上都是可 以检测到是否受到过攻击的水印类型。但是不能通过水印提取来进行所有权的 认证。把脆弱水印添加到联邦学习中，无法实现对保密性攻击的抵御。

（2）鲁棒数字水印

鲁棒数字水印是一种具有抵御攻击能力的数字水印，在遭受一定程度的攻 击之后，也可以凭借其鲁棒性可以从受攻击图片中提取出水印图像去进行所有 权的验证。鲁棒性水印可以分为空间域和变换域两种算法：

空间域的数字水印方式是早期主要使用的方式之一，主要优点集中在易操 作，易实现等方面。空间域的数字水印中最常使用的是最低有效位（LSB）算法， 它的工作原理是通过 LSB 来实现水印的嵌入，这种算法非常容易实现，但是相 对的是他的抗攻击能力很弱，在遭受到攻击之后，并不能完整的从图像中提取 出水印图像，所以说这种算法的鲁棒性不高。

变换域的数字水印方式是通过对变换域系数修改而产生的。Cox[52]提出了一 种基于 DCT 变换的扩频水印技术。该算法通过对肉眼观测的优先性进行研究， 目的是在保证鲁棒性的同时也要有不错的不可见行。算法的原理时将图像进行 离散余弦变换，只保留分流部分，同时把数值大的系数进行修改，最终实现水 印的嵌入过程。

Piva[53]提出了一种基于 DCT 变换的扩频水印技术。通过对载体图像进行离

13

中国知网 https:Www  cnknet

散余弦变换把水印图像嵌入到 DCT 域的中频区域。

Hsu 和 Wu[54]将图像进行 8×8 分块，选择一个二进制序列作为水印信息嵌入 到 DCT 的中频区。还有一种方法就是把高频系数作为水印的嵌入位置。

Zeng[55]等基于上述考虑，提出先修改 DCT 变换的低频系数，然后在中频系 数中嵌入水印信息。

Wang[56]提出了在视觉重要的小波系数中嵌入水印的方法。这种算法根据肉 眼的观测习惯对 DWT 的系数进行操作从而实现水印的嵌入。该算法在嵌入水印 图像时有两种嵌入方案，可以自由选择是否需要原始图像。这种算法有很强的 适用性，即使没有原始图像的信息，也可以实现水印的嵌入，相对的是，它的 鲁棒性并不高。

Kundue[57]提出了基于小波变换的私有水印算法和公开水印算法。私有水印 算法是通过 DWT 得到的，首先将载体图像和水印图像分别进行 DWT 之后，对得 到的子图像进行水印的嵌入，这种算法在水印的提取阶段需要使用到经过 DWT 得到的子图像才能提取出水印图像；公开水印算法对 DWT 进行量化从而实现水 印的嵌入，与之相对的，这种算法在水印的提取阶段不需要其他条件就可以提 取出水印图像。

Chen[58]提出的 QIM 技术，这种方法是根据水印信息不同将信号调制到不同 区间，提取时根据图像像素或变换域系数所落入区间判定提取水印。

Cox[59]提出的扩频调制是利用数字水印与信号处理技术的共同点，降低了以 图像视觉感知重要部分作为嵌入位置时产生的高鲁棒性与低不可见性之间的冲

突[60-61]。

**2.2.3 离散小波变换**

离散小波变换（DWT）是 JPEG-2000 静态数字图像压缩编码标准的核心技术。 图像经过 DWT 变换后会被分解成四个子图像，每个图象都有各自的特征，因此 可以根据实际情况在不同的子图像上嵌入水印图像，这就是使用 DWT 实现数字 水印的方法。在子图像上嵌入水印能够防止图像在经历压缩后导致水印图像的 丢失，同时根据四个子图像，可以调整水印的位置，实现水印的不可见行和鲁 棒性的选择。同时，图像经过 DWT 得到的子图像与人眼视觉模型（HVM）相吻合，

14

中国知网 https:Www  cnknet

有助于更合理地确定水印信息的嵌入位置。

DWT 和 IDWT 的描述为如公式 2.10 和 2.11 所示：

H(w) = Σkhk. e-jkw ................... (2.10)

H(w) = Σk Gk. e-jkw ................... (2.11)

H(w)和H(w)满足正交关系如公式 2.12 所示：

I H(w) I2 +I G(w) I2 = 1 ................ (2.12)

例如，如下给出的H(w)和G(w)是典型的 Haar 小波滤波器

 e-jw .................... 

 一  e-jw .................... 

其中j = J + 1,J, … ,J0,当fj+1(K) = F(f),K∈Z ; J + 1是分辨率最高尺度指 数, J0 是分辨率最低尺度数, F(n)的 DWT 系数：

fJow (k), fJigh(k), fJht(k), … , fJhigh(k) ........... (2.15)

其中fJow (k)是F(n)的低频子带部分，fJhigh(k)是F(n)的其他部分（LH、HL

和 HH）。图像的 DWT 的迭代过程：



以 200×200 大小的 Lena 为图像为例， 图 2.6 展示了图像经过离散小 波变换（DWT）后的结果：

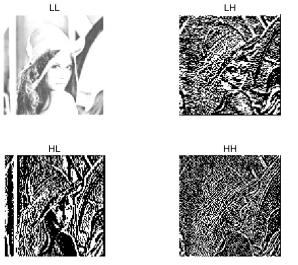


图 2.6 Lena 图像及图像的离散小波变换

15

中国知网 https:Www  cnknet

图像经过 DWT 后的四条子带如图 3.1 右侧所示，与图 3.1 左侧的 Lena 图

像相比，绝大部分能量集中在低频子带 LL，因此其具有较强的抗干扰性和稳定 性。垂直方向的细节部分 LH、水平方向细节部分 HL 和对角线方向的细节部分、 HH 则展现了原始图像的边缘、轮廓和纹理细节等特征。

**2.2.4 数字水印的评价指标**

数字图像水印的基本设计要求包括：

(1)不可见性：嵌入水印后的图像对人眼来说应不可察觉地发生变化，理想 情况下，含水印图像与原始图像在视觉上应完全相同。这是水印算法的基础要 求。

(2)鲁棒性：数字水印应具有一定的持久性，不易被消除。

(3)安全性；数字水印应能承受各种故意攻击，且难以被他人复制或伪造。

(4)效率性：水印提取过程应高效，且提取的水印应能清晰地标识版权所有 者。

(5)抵抗变化：与鲁棒性不同，这一要求强调水印一旦嵌入，攻击者难以修 改或伪造。

水印算法通常使用峰值信噪比(PSNR)和结构相似度(SSIM) 来评估其不可见 性；

水印算法通常使用归一化相关系数(NC)来衡量原始水印图像与提取水印图 像的相似度从而评估其鲁棒性。

（1）PSNR

峰值信噪比（PSNR）以分贝（dB）为单位，其值越大表示图像质量越高， 即两幅图像的可视差异越小。PSNR的定义如下：

PSNR = 10log10  ................... 

MSE = 2 ......... 

这里， M和N表示图像的高度和宽度，I(m, n) 表示原图像中的像素值, I ∗ (m, n)表示含水印图像中的像素值，MSE为均方误差。

（2）SSIM

结构相似度（SSIM）是衡量两幅图像相似度的另一种客观指标，SSIM取值

16

中国知网 https:Www  cnknet

范围为[0,1]，接近 1 表示图像质量更好，等于 1 时表示两幅图像完全相同。图 像x和图像y之间的SSIM 计算公式如下：

SSIM  ............ 

c1 = (k1 L)2 ....................... (2.20)

c2 = (k2 L)2 ....................... (2.21)

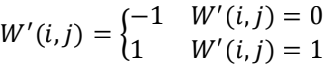
其中μx 是图像x所对应的像素的平均值，μy 是图像y所对应的像素的平均值， σx 是图像x所对应的像素的方差，σy 是图像y所对应的像素的方差，σxy 是图像x 和图像y所对应的像素的协方差，c1 和c2 是常数，L是像素的动态范围，k1 取值 0.01，k2 取值 0.03。

（3）NC

归一化相关系数（NC）是一个客观指标，用于衡量图像之间的相似性。对 于数字水印的确权有两种方式：一种是不提取水印直接检测数字水印来进行确 权的水印检测法，另一种是通过提取出的水印进行确权的水印提取法。水印检 测法是检测图像中是否包含水印，不需要对图像进行水印的提取。而水印提取 法则是从水印图像中提取出水印图像，同时与原始图像进行对比来进行所有权 的验证。那么提取的水印和原始水印的相似程度的判断就很重要，肉眼对于两 张相似图片的判断主观性的影响过大，因此需要一个可靠的评价指标，归一化 相关系数（NC）。NC取值范围为[0,1]，其中NC越接近 1 表示图像相似度越高， 当 NC 等于 1 时，表示两幅图像完全一致。其定义如下：

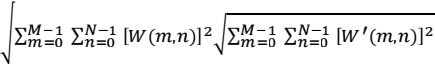
设W为原始水印，W ′ 为提取出的水印。对W和W′ 进行以下变换：

 ................ 

 ................ 

那么归一化相关系数(NC)可以表示为：

 ......... 



其中, M和N表示水印图像的长度和宽度, W(m, n)表示原始水印图像中的 像 素 点 位 置 , W ′ (m, n) 表 示 提 取 出 的 水 印 图 像 中 的 像 素 点 的 位 置 。

17

中国知网 https:Www  cnknet

**2.3 本章小结**

本章主要介绍了隐私计算的四种实现方式，并重点讨论了联邦学习的基本 原理和实现方法。同时，介绍了数字水印的基本原理以及一些图像质量评价指 标和方法，用于验证本文提出算法的有效性。

18

中国知网 https:Www  cnknet

**第** **3 章基于** **DWT-SVD** **的双重鲁棒数字水印算法**

在联邦学习的背景下，保密性攻击构成了一个显著的安全挑战。为了提升 联邦学习的安全性，一种可行的方法是在学习开始前对图像进行数字水印处理。 这样，即使遭受保密性攻击，数字水印也能助于确认图像的所有权。

在此背景下，数字水印应满足两个关键要求：不可见性和提取准确性。

不可见性：水印的不可见性应当强， 以确保加水印的图像不会影响联邦学 习的效果。这可以通过峰值信噪比（PSNR）和结构相似度（SSIM）来评估。高 PSNR 和 SSIM 值意味着更好的不可见性，从而保证联邦学习中使用的数字水印 不会影响学习结果。

提取准确性：提取水印时，应确保提取的水印与原始水印图像高度相似， 以准确地确认水印的归属。这可以通过归一化相关系数（NC）来评估，接近 1 的值表示高度相似，从而有助于准确确定图像的所有权。

为了在联邦学习中提高数字水印技术的效果，技术指标 PSNR、SSIM 和 NC 的要求应尽可能高。鲁棒数字水印凭借其不可见性和鲁棒性，成为这种应用场 景的理想选择。因此，通过分析鲁棒数字水印算法的原理，可以提出一种更有 效的鲁棒数字水印算法。

**3.1 基于奇异值分解（SVD）的鲁棒数字水印算法**

**3.1.1 鲁棒数字水印算法**

鲁棒水印能在遭受攻击后保持不变或仅有微小变化，以证明所有权。数字 图像在互联网中无论是在通信、新闻还是网购中都占有重要地位。数字图像鲁 棒水印技术结合了数字图像特点与信息隐藏技术，为开放网络环境中的图像提 供保护，是数字水印技术的一个重要分支。接下来，通过介绍奇异值分解（SVD） 的原理，可以实现基于 SVD 的鲁棒数字水印算法。

19

中国知网 https:Www  cnknet

**3.1.2 奇异值分解**

奇异值分解（SVD）在数学领域广泛使用，并可以应用在很多的领域。它作 为矩阵分解的一种方法，可以将一个矩阵分解到的三个矩阵，通过对分解得到 的矩阵进行分析，可以很好的了解矩阵的相关性。由于这个特性，奇异值分解 也可以应用在图像领域，实现对图像的分析。

对于一张图像，如果对其进行图片的修改操作，那么图像对应的奇异值矩 阵也会随之发生变化，这种变化是根据图像的变化而改变的，但奇异值矩阵的 变换是微小的。凭借这种特性，可以很好的应用在数字水印中，一旦图片遭受 攻击，通过奇异值的变化就可以反映出来，而微笑的改变也更容易提取出准确 的嵌入的数字水印，极大的提升了数字水印的鲁棒性。

SVD 分解一个图像会得到三个矩阵：左奇异矩阵 U、奇异值矩阵 S 和右奇异 矩阵 V。对于 M×N 的矩阵 P，它的奇异值分解可以表示为：

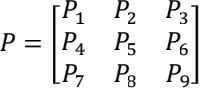
P = U ∗ S ∗ VT ....................... (3.1)

左奇异矩阵U是 M×M 阶的酉矩阵，右奇异矩阵V是 N×N 阶的酉矩阵，VT 是 V的共轭转置。奇异值矩阵S是 M×N 阶的半正定对角矩阵，它的对角线元素即 为矩阵P的奇异值。

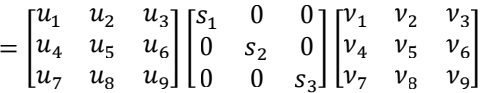
奇异值S1 ≥ S2 ≥ S3 … ≥ SN≥0 ，由此S被矩阵P唯一确定。公式(3.2)描述了 SVD 分解后的矩阵重构过程。

U ∗ S ∗ VT = P ....................... (3.2)

假设 P 是 3×3 的矩阵，则 SVD 算法中矩阵的表示公式为：



........ (3.3)



**3.1.3 基于奇异值分解的鲁棒数字水印算法**

通过对奇异值分解原理的分析，可以知道对于图像，它的奇异值是很重要的 一张特征。并且对于图片的修改，奇异值的变化的程度很小，保证了它的稳定 性。如果将水印添加在奇异值上，图像遭受了攻击，可以凭借奇异值的稳定性

20

中国知网 https:Www  cnknet

提取出水印经行所有权的验证，证明这种水印嵌入的方法具有一定的鲁棒性。 假设有一个矩阵 M×N 的矩阵A ，则A对应的SVD矩阵分解表示为:

A = U \* S \* VT ....................... (3.4)

SVD 分解一个图像会得到三个矩阵：左奇异矩阵 U、奇异值矩阵 S 和右奇异 矩阵 V。奇异值矩阵 S 具体描述如下：

S = diag(σ1, σ2, … , σm ), σ1 ≥ σ2 ≥ … ≥ σm ≥ 0 ......... (3.5)

对水印图像I进行SVD可以得到UI ，SI 和VI，对载体图像W进行SVD可以得到 UW , SW 和VW。对水印图像I进行SVD可以得到UI ，SI 和VI 依次加到载体图像W 进行SVD可以得到UW , SW 和VW：

W = UW \* SW \* V *，*I = UI \* SI \* VIT ............. (3.6)

将水印通过 SVD 实现嵌入，得到嵌入后的水印图像S：

S = SW + λSI ....................... (3.7)

其中,0< λ <1 代表嵌入系数。为了选择合适的嵌入系数λ , 考虑到隐藏性 和鲁棒性的影响，采用以下方法确定λ:

 .................... 

其中，PSNR是峰值信噪比，描述的是原始载体图像和嵌入水印图像的相似 程度；NC是相关系数，描述的是原始水印图像和提取水印图像的相似性。

载体图像I选择的是 200×200 的 Lena 图像，水印图像W选择的是 100×

100 的二值图像。以嵌入水印奇异值矩阵S方法为例，介绍水印嵌入和水印提取 过程，实现的结构图如图 3.1 和图 3.2 所示：

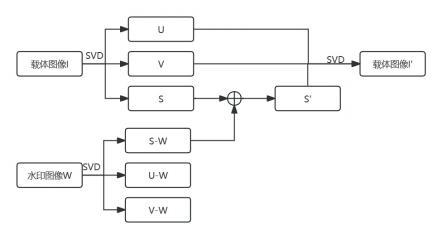


图 3.1 基于奇异值分解的鲁棒数字水印嵌入结构图

21

中国知网 https:Www  cnknet

根据图 3.1，水印嵌入的过程可分为以下几个步骤：

（1）对原载体图像I进行奇异值分解（SVD）：I = U\*S\*V。

（2）对水印图像W进行奇异值分解（SVD）：W = U-W ∗ S-W ∗ V-W。

（3）设置嵌入量af = 0.1 ，分别将水印图像W的奇异值S-W加到原始载体 图像I的奇异值S上得到：S′ = S + af ∗ S-W ，进而得到水印嵌入后的载体图像I′： I ′ = U\*S′\*V。

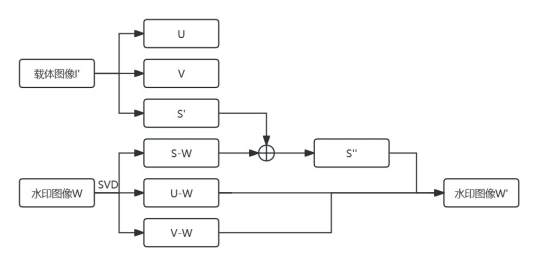


图 3.2 基于奇异值分解的鲁棒数字水印提取结构图 根据图 3.2，水印提取的过程可分为以下几个步骤：

（1）对嵌入水印后的载体图像I′ 进行奇异值分解（SVD）处理得到：I ′ = U\*S′\*V。

（2）进而得到提取水印图的奇异值矩阵S′′ ：S′′ = (S′ − S-W)/af。

（3）将得到提取水印图的奇异值矩阵S′′ 与水印图像W的奇异值矩阵U-W 和V-W相结合得到提取出的水印图像W1：W1 = U-W ∗ S′′ ∗ V-W。

通过实验实现，得到的结果如图 3.3 所示， 同时计算载体图像与嵌入水 印图的 PSNR 和SSIM 值，水印图像与提取水印图像的 NC 值，如表 3.1 所示。

表 3.1 基于 SVD 鲁棒数字水印算法的 PSNR、SSIM、NC

|  |  |  |  |
| --- | --- | --- | --- |
|  | S | U | V |
| PSNR | 20.6093 | 25.5324 | 25.4580 |
| SSIM | 0.9793 | 0.9996 | 0.9633 |
| NC | 1.0000 | 0.0932 | 0.1599 |

22

中国知网 https:Www  cnknet

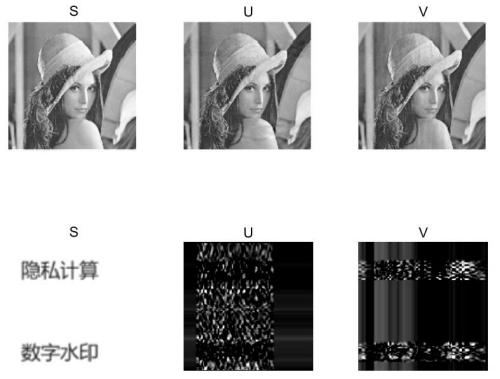


图 3.3 基于奇异值分解的鲁棒数字水印算法

同时通过上述实验可以得出，在水印的质量上奇异值 S 具有很高的性能， 在嵌入水印的载体图像的图像质量上奇异值 U、V 具有更好的性能。为了提升数 字水印的不可见性和鲁棒性，将在嵌入水印前使用 DWT（离散小波变换）进行 优化。

**3.2 基于** **DWT-SVD 的鲁棒水印算法**

3.1 节提出基于奇异值分解的鲁棒数字水印算法的鲁棒性很好，但是不可 见性略微偏低。在本节中通过使用引用 DWT（离散小波变换） 来进行优化。并 分析 Anand[62]提出的一种 DWT-SVD 水印技术，通过调整得到本算法。

**3.2.1 单奇异值** **S 水印算法**

载体图像I选择的是 200×200 的 Lena 图像，水印图像w选择的是 100×

100 的二值图像。以嵌入水印奇异值矩阵S加到载体图像的LL分量上为例，介绍 水印嵌入和水印提取过程，实现的结构图如图 3.4 和图 3.5 所示。

23

中国知网 https:Www  cnknet

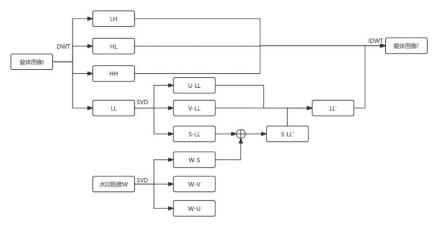


图 3.4 基于 DWT-SVD 的单奇异值水印嵌入结构图 根据图 3.4，水印嵌入的过程可分为以下几个步骤：

（1 ）首先对载体图像I进行离散小波(DWT)变换， 可以的到四个部分 （LL,LH,HL, HH） :I = DWT(LL,LH,HL, HH)。

（ 2 ）将 LL 子 带 进 行 奇 异 值 分 解（ SVD ） 处 理 得 到 奇 异 值 ：LL = U-LL\*S-LL\*V-LL。

（3）对水印图像W进行奇异值分解（SVD）处理得到奇异值：W = U-W ∗ S-W ∗ V-W。

（4）设置嵌入量af = 0.1，将加密后的水印图像W′ 的奇异值S-W加到奇异 值S-LL上得到：S-LL′ = S-LL + af ∗ S-W ，进而得到：LL′ = U-LL\*S-LL′ \*V-LL。

（5）将上一步得到的LL′ 与载体图像I经过DWT得到的LH 、HL和HH进行离 散小波变换逆变换（IDWT），可以实现数字水印的嵌入， 得到水印嵌入后的载 体图像I′ ：I ′ = IDWT(LL′ ,LH ,HL, HH)

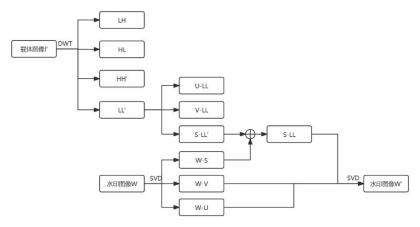


图 3.5 基于 DWT-SVD 的单奇异值水印提取结构图

24

中国知网 https:Www  cnknet

根据图 3.5，水印提取的过程可分为以下几个步骤：

（ 1 ）对 嵌入水 印后 的载体 图像I′ 进行 离散 小波变换（ DWT ）：I ′ = IDWT(LL,LH,HL, HH)。

（2）将LL子带进行奇异值分解（SVD）处理得到：LL = U\*S\*V ，进而得到 提取水印图的奇异值矩阵：S ′ = (S − S-W)/af。

（3）将上一步得到的奇异值矩阵S′ 与对水印图像W经过SVD得到奇异值 U-W和V-W相结合得到提取出的水印图像W1：W1 = U-W ∗ S ′ ∗ V-W

表 3.2 基于 DWT-SVD 的单奇异值 S 水印算法的 PSNR、SSIM、NC

|  |  |  |  |
| --- | --- | --- | --- |
|  | PSNR | SSIM | NC |
| LL | 23.3321 | 0.9995 | 1.0000 |
| LH | 28.0001 | 0.9999 | 0.9980 |
| HL | 27.9064 | 0.9999 | 0.9602 |
| HH | 28.0108 | 0.9999 | 0.9991 |

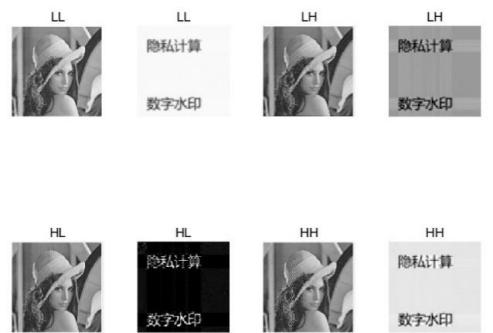


图 3.6 基于 DWT-SVD 的单奇异值 S 水印算法

这种情况下，通过分析数字水印的评价指标发现，无论是嵌入水印的载体 图像还是提取的水印图像，都很优秀，证明这种算法的实用性很高。

**3.2.2 单奇异值** **U 水印算法**

将水印图像经 SVD 分解得到的奇异值 U 分别加到原始图像 DWT 后的四个部 分的 SVD 分解后的奇异值 U 上。得到的嵌入水印的载体图像和数字水印的评价 指标如表 3.3 和图 3.7 所示：

25

中国知网 https:Www  cnknet

表 3.3 基于 DWT-SVD 的单奇异值 U 水印算法的 PSNR、SSIM、NC

|  |  |  |  |
| --- | --- | --- | --- |
|  | PSNR | SSIM | NC |
| LL | 23.3266 | 0.9995 | 0.0936 |
| LH | 27.9980 | 0.9999 | 0.4301 |
| HL | 28.0034 | 0.9999 | 0.0944 |
| HH | 28.0102 | 0.9999 | 0.1270 |

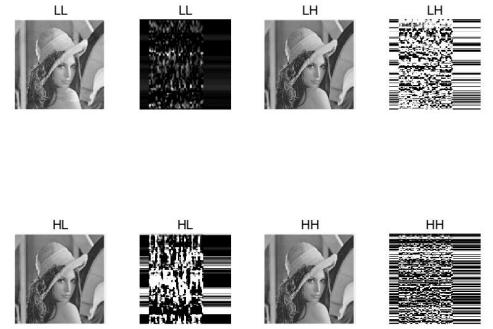


图 3.7 基于 DWT-SVD 的单奇异值 U 水印算法

这种情况下，通过分析数字水印的评价指标发现，嵌入水印的载体图像质 量很高，说明可以奇异值 U 提升 PSNR 和SSIM，从而提升图片质量。

**3.2.3 单奇异值** **V 水印算法**

将水印图像经 SVD 分解得到的奇异值 V 分别加到原始图像 DWT 后的四个部 分的 SVD 分解后的奇异值 V 上。得到的嵌入水印的载体图像和数字水印的评价 指标如表 3.4 和图 3.8 所示：

表 3.4 基于 DWT-SVD 的单奇异值 V 水印算法的 PSNR、SSIM、NC

|  |  |  |  |
| --- | --- | --- | --- |
|  | PSNR | SSIM | NC |
| LL | 23.3266 | 0.9995 | 0.1279 |
| LH | 27.9980 | 0.9999 | 0.1142 |
| HL | 28.0034 | 0.9999 | 0.4426 |
| HH | 28.0102 | 0.9999 | 0.4166 |

26

中国知网 https:Www  cnknet

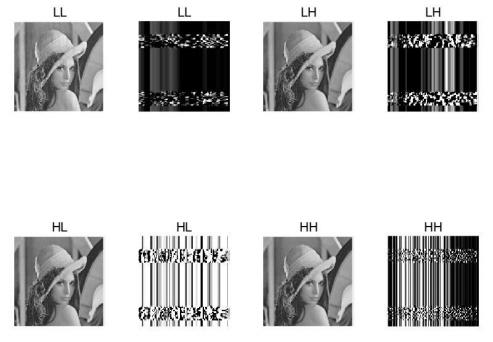


图 3.8 基于 DWT-SVD 的单奇异值 V 水印算法

这种情况下，通过分析数字水印的评价指标发现，嵌入水印的载体图像质 量很高，但说明可以奇异值 V 提升 PSNR 和SSIM，从而提升图片质量。

通过分析三种情况下的 PSNR、SSIM 和 NC 的值，可以得出结论，通过把水 印图像的 S 加到原始图像经过 DWT 后的 LL 分量上水印提取质量最高；通过把水 印图像的 U、V 加到原始图像经过 DWT 后的 LH、HL 和HH 分量上的嵌入水印的载 体图像质量高。

**3. 3 基于** **DWT-SVD 的双重奇异值水印**

上述的数字水印算法，单奇异值 S 嵌入水印算法的提取的水印图像质量很 高，但是与单奇异值 U 嵌入水印、单奇异值 V 嵌入水印算法相比，其 PSNR 和 SSIM 值略低。因此尝试将奇异值 S 与奇异值 U 和 V 相结合，在保证 NC 值不变 的情况下，提高 PSNR、SSIM 的值。

27

中国知网 https:Www  cnknet

**3.3.1 双重奇异值** **S 、U 水印算法**

载体图像 I 选择的是 200 × 200 的 Lena 图像，水印图像 W 选择的是 100 × 100 的二值图像。 以将载体图像进行离散小波变换（DWT）得到的四 个部分，分别和水印图像进行 SVD 分解得到的奇异值矩阵S和V相结合嵌入水 印为例，介绍水印嵌入和水印提取过程，实现的结构图如图 3.9 和图 3.10 所示。

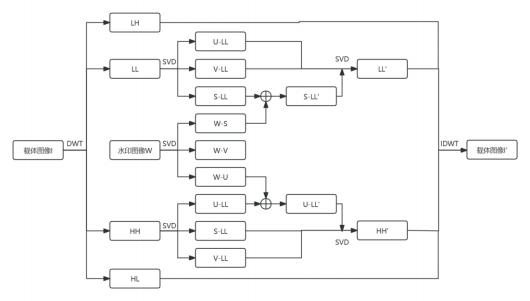


图 3.9 基于 DWT-SVD 的双重奇异值水印的嵌入结构图 根据图 3.9，水印嵌入的过程可分为以下几个步骤：

（1）首先对载体图像I进行离散小波(DWT)变换，可以的到四个部分:I = DWT(LL,LH,HL, HH)。

（2）对 LL 子带进行奇异值分解（SVD）得到：LL = U-LL\*S-LL\*V-LL。

（3）对水印图像W进行SVD得到：W = U-W ∗ S-W ∗ V-W。

（4）设置嵌入量af = 0.1，将水印图像W的奇异值S-W加到奇异值S-LL上 得到：S-LL′ = S-LL + af ∗ S-W，进而得到：LL′ = U-LL\*S-LL′ \*V-LL。

（5）将水印图像W′ 的奇异值矩阵V-W加到奇异值矩阵V-LH 上(V-LH 、 V-HL和V-HH三者随机一个，这里以V-LH为例)上得到 V- V-LH′ ：V-LH′ = V-LH + af ∗ V-W，进而得到：LH′ = U-LH\*S-LH\*V-LH′。

（6）将上述得到的LL′ 、LH′ 、HL和HH通过进行离散小波变换的逆变换 （IDWT）即可得到水印嵌入后的载体图像I′ ：I ′ = IDWT(LL′ ,LH′ ,HL, HH)

28

中国知网 https:Www  cnknet

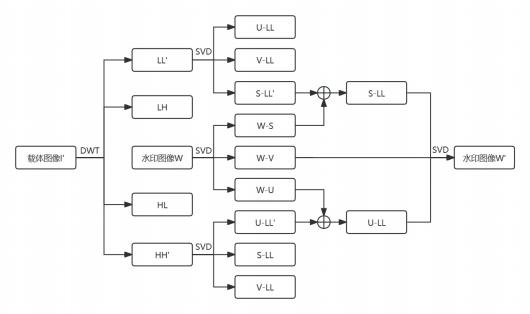


图 3.10 基于 DWT-SVD 的双重奇异值水印的提取结构图 根据图 3.10，水印嵌入的过程可分为以下几个步骤：

（1） 对嵌入水印后的载体图像I′ 进行离散小波变换（DWT）： I ′ = IDWT(LL,LH,HL, HH)

（2）将LL子带进行奇异值分解（SVD）处理得到：LL = U\*S\*V，进而得到 提取水印图的奇异值矩阵 ：S ′ = (S − S-W)/af； 将LH子带进行奇异值分解 （SVD）处理得到：LH = U\*S\*V ，得到提取水印图的奇异值矩阵：V ′ = (V − V-W)/af。

（3）将提取水印图像的奇异值矩阵S′ 与原始水印图像W′ 的奇异值矩阵U-W 和V-W相结合得到提取出的水印图像W1： W1 = U-W ∗ S ′ ∗ V-W

将水印图像的 S 和 U 分别加到原始图像 DWT 后的四个部分中的两种上，得 到的嵌入水印的载体图像和数字水印的评价指标如表 3.5 和图 3.11 所示：

表 3.5 双重奇异值 S、U 水印算法的 PSNR、SSIM、NC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PSNR | SSIM |  | PSNR | SSIM |
| LL-LH | 23.9042 | 0.9996 | HL-LL | 22.0524 | 0.9995 |
| LL-HL | 23.8656 | 0.9996 | HL-LH | 24.5758 | 0.9999 |
| LL-HH | 23.9068 | 0.9996 | HL-HH | 24.5825 | 0.9999 |
| LH-LL | 22.6606 | 0.9995 | HH-LL | 22.1111 | 0.9995 |
| LH-HL | 25.8443 | 0.9996 | HH-LH | 25.1288 | 0.9999 |
| LH-HH | 25.8443 | 0.9999 | HH-HL | 25.0791 | 0.9999 |

29

中国知网 https:Www  cnknet

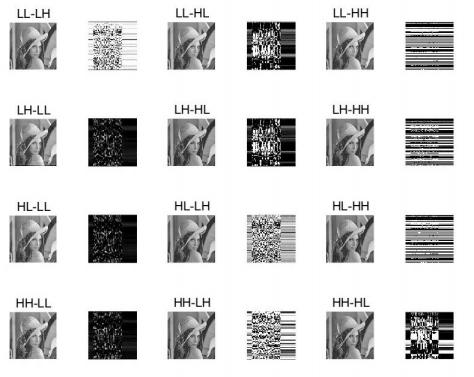


图 3.11 双重奇异值 S、U 水印算法

**3.3.2 双重奇异值** **S 、V 水印算法**

将水印图像的 S 和 V 分别加到原始图像 DWT 后的四个部分中的两种上，得 到的嵌入水印的载体图像和数字水印的评价指标如图 3.12 和表 3.6 所示：

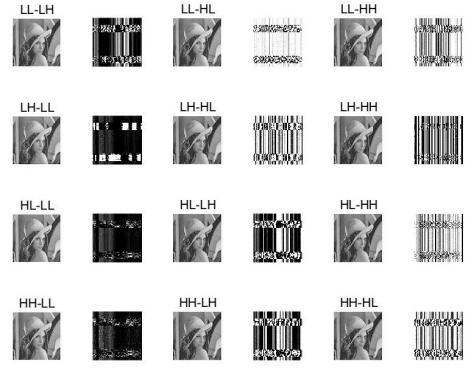


图 3.12 双重奇异值 S、V 水印算法 30

中国知网 https:Www  cnknet

表 3.6 双重奇异值 S、U 水印算法的 PSNR、SSIM、NC

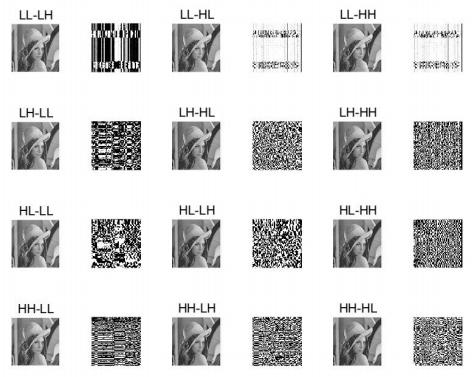
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PSNR | SSIM |  | PSNR | SSIM |
| LL-LH | 25.8614 | 0.9998 | HL-LL | 23.9687 | 0.9998 |
| LL-HL | 25.8596 | 0.9998 | HL-LH | 25.2965 | 0.9999 |
| LL-HH | 25.8633 | 0.9998 | HL-HH | 25.2985 | 0.9999 |
| LH-LL | 24.9004 | 0.9998 | HH-LL | 24.7485 | 0.9998 |
| LH-HL | 26.5790 | 0.9996 | HH-LH | 26.6158 | 0.9999 |
| LH-HH | 26.5790 | 0.9999 | HH-HL | 26.6147 | 0.9999 |

**3.3.3 双重奇异值** **U 、V 水印算法**

将水印图像的 U 和 V 分别加到原始图像 DWT 后的四个部分中的两种上，得 到的嵌入水印的载体图像和数字水印的评价指标如表 3.7 和图 3.13 所示：

表 3.7 双重奇异值 U、V 水印算法的 PSNR、SSIM、NC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | PSNR | SSIM |  | PSNR | SSIM |
| LL-LH | 25.4966 | 0.9998 | HL-LL | 25.5756 | 0.9998 |
| LL-HL | 25.4992 | 0.9998 | HL-LH | 27.9492 | 0.9999 |
| LL-HH | 25.4990 | 0.9998 | HL-HH | 27.9533 | 0.9999 |
| LH-LL | 25.6098 | 0.9998 | HH-LL | 25.6122 | 0.9998 |
| LH-HL | 28.0038 | 0.9996 | HH-LH | 28.0073 | 0.9999 |
| LH-HH | 28.0038 | 0.9999 | HH-HL | 28.0090 | 0.9999 |



`

图 3.13 双重奇异值 U、V 水印算法

上述的情况中，加密图像的 PSNR 和 SSIM 值有明显提升，说明水印的嵌入

质量变高。与之前的 SVD 得到的数据相对比，可以得出结论，在水印的提取阶 31

中国知网 https:Www  cnknet

段，水印图像通过 SVD 得到的 S 分量起到主要作用，在解密水印时得到一个与 水印图像相近的解密水印图像；在水印的嵌入阶段，水印图像通过 SVD 得到的 V 分量起到主要作用，在水印嵌入时，可以得到一个与原始图像相近的嵌入水 印的图像。

为了提升嵌入水印图像的质量（PSNR 值和SSIM 值），接下来在水印的嵌入 阶段时，把水印图像经过 SVD 后的 S 分量加到原始图像通过 DWT 后的LL 分量上， 把水印图像经过 SVD 后的 V 分量加到原始图像通过 DWT 后的 LH、HL、HH 分量上； 在水印的提取阶段，只将嵌入水印图像的 LL 分量提取出S 分量用于水印图像的 提取过程，得到的结果如表 3.8 和图 3.14 所示

表 3.8 改进的双重奇异值 S、V 水印算法的 PSNR、SSIM、NC

|  |  |  |  |
| --- | --- | --- | --- |
|  | PSNR | SSIM | NC |
| LL-LH | 25.8614 | 0.9998 | 1.0000 |
| LL-HL | 25.8596 | 0.9998 | 1.0000 |
| LL-HH | 25.8633 | 0.9998 | 1.0000 |



图 3.14 改进的双重奇异值 S、V 水印算法

这种情况下，较之前的数字水印方案，性能有明显的提升。而这三种数字 水印的方式的性能都很高，因此提出一种基于 DWT-SVD 的双重鲁棒数字水印算 法，具体实现方式是通过把水印图像经过 SVD 变换得到的 S 分量，加到原始图 像通过 DWT 得到的 LL 分量经过 SVD 变换的S 份量上；再将通过把水印图像经过 32

中国知网 https:Www  cnknet

SVD 变换得到的 V 分量，加到原始图像通过 DWT 得到的 LH、HL、HH 分量中随机 的一种分量经过 SVD 变换的 V 份量上。

**3.4 水印预处理**

为了进一步提升数字水印的安全性，在嵌入水印前，对水印进行预处理， 即进行一次加密，这里选取的是 Arnold 变换，接下来介绍 Arnold 变换的原理 以及用实验来说明应用在数字水印中提升安全性的效果。

**3.4.1 Arnold 变换**

Arnold 变换，也被称为猫脸变换，是一种矩阵加密的方法。它的工作原理 对矩阵进行变换，这种变换的次数是有限的。凭借这种变换，可以将矩阵变为 乱序状态，可以实现矩阵的加密。

当给定图像f(x, y)，通过使用二维 Arnold 变换，可以改变矩阵内的数据， 把图像当成一个矩阵，对于图像来说就是改变图像的像素值。这样就会导致图 像的像素值变得混乱无序，从而实现图像加密的功能。

对于一幅大小为 N×N 的图像，像素点(xn, yn )通过可以 Arnold 变换成 (xn+1, yn+1)，这种变换可以通过公式（3.10）进行：这种变换是对图像的像素 点进行位置的变换，将整幅图像的像素点全都进行改变，也就是说在经过足够 次数的 Arnold 变换之后，像素值变得混乱无序的图像也会恢复成原来的图像， 说明 Arnold 变换是具有周期性的。

[] = [ ab  1] [] (modN) = A [] (modN) ....（3.9）

其中的mod(N)是对N取余。这么做的目的是保证经过变换后的像素点的变 化都实在图像中进行的。理想的情况是图像的长和宽是一样的，这样无论怎样 进行变化都是发生在图像的内部。

Arnold 变换也有逆变换方法，需要事先知道参数 a、b 的值，具体的逆变 换由公式(3.101)所示:

[] = [a 1 b] [] mod(N) ..........（3.10）

33

中国知网 https:Www  cnknet

当 a=2、b=2 时，Arnold 变换和逆变换公式如下所示：

 ............. 

[] = 2 2] [] mod(N) ............（3.12）

**3.4.2 水印预处理实验仿真**

使用 100 × 100 大小的水印图片进行 Arnold 变换的效果及其直方图统计 如图 3.15 和图 3.16 所示。

图 3.10 是在 n=10、n=20、n=50、n=150 时的置乱图像和置乱还原图像。

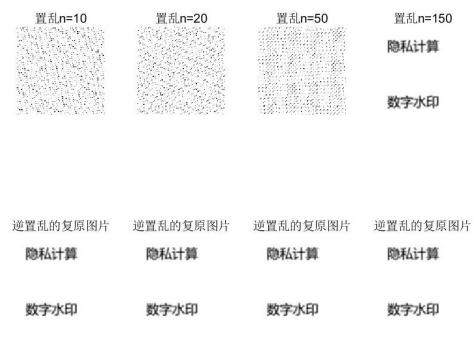


图 3.15 水印置乱图像与复原图像

从 Arnold 变换图像来看，当 n=150 时，Arnold 变换图像和原始图像一致， 说明 Arnold 变换具有周期性。利用周期性，当矩阵进行一定次数变换的时候，

就会恢复原图。当 a=1、b=1 时，探究当 N 取不同值时，Arnold 变换图像与原 始图像一致的次数如下表 3.9 所示。

表 3.9 恢复原始图像的 Arnold 变换次数

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | 4 | 32 | 50 | 64 | 100 | 128 |
| 变换次数 | 3 | 24 | 150 | 59 | 150 | 96 |

原始图像以及在 n=10、n=20、n=50、n=150 时置乱图像的直方图如图 3.17 所示。

34

中国知网 https:Www  cnknet

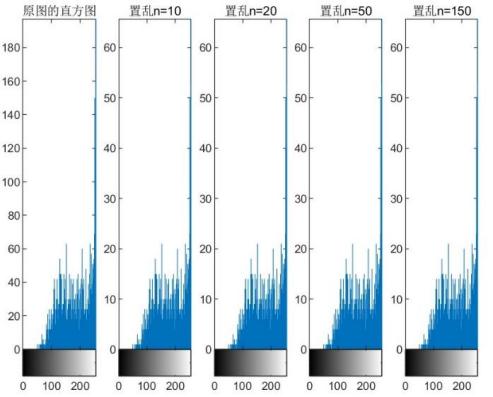


图 3.16 原始图像和 Arnold 变换后图像的直方图

通过图 3.16 所示，原始图像和 Arnold 变换后图像的直方图没有不同，这 正是 Arnold 变换的功能体现，它仅仅改变了图像中像素点的位置，没有去对像 素值进行修改，从而不会影响图像的直方图。

从上述的实验可以得出，当有一张经过 Arnold 变换的图像想要恢复原始图 片时，有两种办法：一种是通过逆 Arnold 变换的方式，这种方式需要知道事先 设定的 a、b 值才能复原，而这两个参数的设置是随机的，因此通过破解 a、b 参数这种方式恢复原始图片是很难实现的；另一种方式是通过 Arnold 变换的周 期性来进行的，对图像一直进行 Arnold 变换的变换，直到恢复出原始图像。但 这种方式依旧需要事先知道 a、b 的参数，因此这种方式也很难恢复出原始图像。 综上所述，Arnold 变换具有很高的保密性。

图像通过 Arnold 变换后具有很强的保密性，可以应用到在 3.3 节提出的基 于脆弱数字水印和鲁棒性数字水印的双重水印方式中去，在嵌入水印图像前把 水印图像通过 Arnold 变换，在提取水印图像后再经过逆 Arnold 变换恢复出原 始水印图像。

35

中国知网 https:Www  cnknet

**3.5 实验仿真平台搭建**

实验仿真平台的搭建如表 3.10 所示。在此基础上完成模型的搭建、训练以 及测试任务。

表 3.10 实验仿真平台配置

|  |  |
| --- | --- |
| 环境配置 | 参数配置 |
| CPU GPU  开发语言  集成开发软件 | Ryzen9 5900HX  RTX 3080  MATLAB 脚本语言  MATLAB R2023a |

**3.6 基于** **DWT-SVD 的双重鲁棒数字水印的算法实现**

在 3.5 节中得知Arnold 变换可以提升图片的安全性，将其应用在数字水印 中也可以进一步提升数字水印的安全性。载体图像I选择的是 200×200 的 Lena 图像，水印图像w选择的是 100×100 的二值图像。通过加入 Arnold 变换，进 一步完善基于 DWT-SVD 的双重鲁棒数字水印的算法。

**3.6.1 水印的嵌入**

介绍基于 DWT-SVD 的双重鲁棒数字水印的算法水印嵌入过程，实现的结构 图如图 3.17 所示：

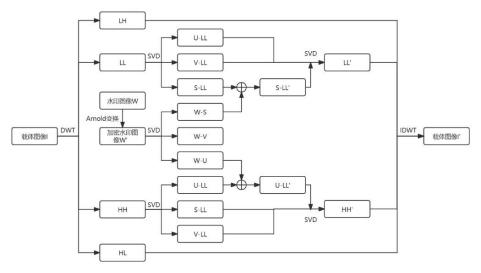


图 3.17 基于 DWT-SVD 的双重鲁棒数字水印的水印嵌入过程 36

中国知网 https:Www  cnknet

根据图 3.17 所示，水印嵌入的过程可分为以下几个步骤：

（1）在预处理阶段，利用 Arnold 变换对大小为 100×100 的水印图像W进 行预处理，参数设置a = 1 、b = 1 、n = 10经过 Arnold 变换得到加密的水印图 像W′。

（2）对原始载体图像 I 进行离散小波(DWT)变换:

I = DWT(LL,LH,HL, HH) ................ (3.12)

（3）将 LL 子带进行奇异值分解（SVD）处理得到奇异值：

LL = U-LL\*S-LL\*V-LL .................. (3.13)

将LH子带进行奇异值分解（SVD）处理得到奇异值：

LH = U-LH\*S-LH\*V-LH ................. (3.14)

将HL子带进行奇异值分解（SVD）处理得到奇异值：

HL = U-HL\*S-HL\*V-HL ................. (3.15)

将HH子带进行奇异值分解（SVD）处理得到奇异值：

HH = U-HH\*S-HH\*V-HH ................ (3.16)

（4）对加密后的水印图像W′ 进行奇异值分解（SVD）处理得到奇异值：

W ′ = U-W ∗ S-W ∗ V-W ................. (3.17)

（5）设置嵌入量af = 0.1，将加密后的水印图像W′ 的奇异值S-W加到奇异 值S-LL上得到：

S-LL′ = S-LL + af ∗ S-W ................. (3.18)

进而得到：

LL′ = U-LL\*S-LL′ \*V-LL ................. (3.19)

将水印图像W′ 的奇异值V-W加到奇异值V-LH (V-LH 、V-HL和V-HH三者 随机一个，这里以V-LH为例)上得到 V- V-LH′：

V-LH′ = V-LH + af ∗ V-W ................ (3.20)

进而得到：

LH′ = U-LH\*S-LH\*V-LH′ ................. (3.21)

（6 ）将上述得到的LL′ 、LH′ 、HL 和HH 进行离散小波变换的逆变换 （IDWT）即可得到水印嵌入后的载体图像：

I = IDWT(LL′ ,LH′ ,HL, HH) .............. (3.22)

37

中国知网 https:Www  cnknet

**3.6.2 水印的提取**

介绍基于 DWT-SVD 的双重鲁棒数字水印的算法水印提取过程，实现的结构 图如图 3.18 所示：

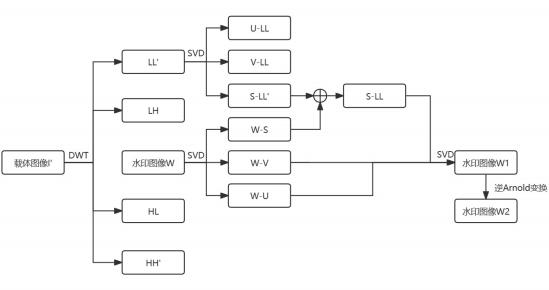


图 3.18 基于 DWT-SVD 的双重鲁棒数字水印的水印嵌入过程 根据图 3.18 所示，水印提取的过程可分为以下几个步骤：

（1）对嵌入水印后的载体图像I′ 进行离散小波变换（DWT）：

I ′ = IDWT(LL,LH,HL, HH) ............... (3.23)

（2）将LL子带进行奇异值分解（SVD）处理得到：

LL = U\*S\*V ....................... (3.24)

进而得到提取水印图的奇异值矩阵：

S ′ = (S − S-W)/af ................... (3.25)

（3）将提取水印图的奇异值矩阵S′ 与加密后的水印图像W′ 的奇异值矩阵 U-W和V-W相结合得到提取出的水印图像W1。

W1 = U-W ∗ S ′ ∗ V-W .................. (3.26)

（4）提取水印图像W1是经过 Arnold 变换后的图像信息，通过进行逆 Arnold 变换得到真正的水印信息W2。

38

中国知网 https:Www  cnknet

**3.6.3 实验仿真**

下图是根据 3.6.1 和 3.6.2 的流程进行实验仿真得到的结果：

表 3.11 基于 DWT-SVD 的双重鲁棒数字水印的 PSNR、SSIM 和 NC

|  |  |  |  |
| --- | --- | --- | --- |
|  | PSNR | SSIM | NC |
| LL-LH | 25.8614 | 0.9998 | 1.0000 |
| LL-HL | 25.8596 | 0.9998 | 1.0000 |
| LL-HH | 25.8633 | 0.9998 | 1.0000 |

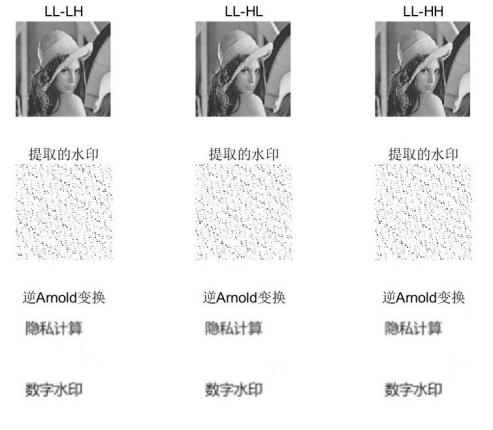


图 3.19 基于 DWT-SVD 的双重鲁棒数字水印算法

根据表 3-11 的数据，这种情况下的数字水印评价指标 PSNR、SSIM 和NC 相 比单独使用 DWT 和 SVD 时更高，说明这种数字水印算法在联邦学习中的更适用。

**3.7 本章小结**

本章提出一种基于 DWT-SVD 的双重鲁棒数字水印算法。首先利用 Arnold 变换对水印图像进行加密，提高了水印图像的安全性，然后将图像进行离散小 波变换，再使用奇异值分解嵌入双重水印，增强水印图像的鲁棒性。实验结果 表明，使用该算法后，保证图像的不可见性的前提下图像的鲁棒性有明显提升。

39

中国知网 https:Www  cnknet

40

中国知网 https:Www  cnknet

**第** **4 章** **利用数字水印增强隐私计算安全性的可行性研究**

在第 3 章提出了适合在联邦学习中使用的数字水印算法后，接下来在第 4 章将这种算法应用到联邦学习中。由于数字水印可能会影响联邦学习的准确率 和时效性，因此需要进行以下两项工作：

1.数据分发和聚合方式分析：针对传统的横向联邦学习方式，分析不同的 数据分发方式和聚合方式，选择准确率更高的方式作为联邦学习的方法。

2.实验验证影响：通过实验验证加入数字水印技术对联邦学习准确率和时 效性的影响。

**4. 1 横向联邦学习模型**

在 2.2 节介绍了联邦学习的基本架构，根据横向联邦学习的特点，下面介 绍一下联邦学习具体的流程图，如图 4.1 所示：

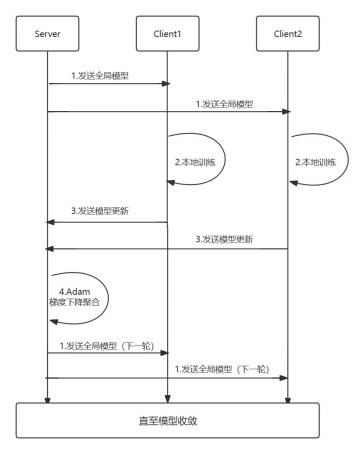


图 4.1 联邦学习流程图

41

中国知网 https:Www  cnknet

联邦学习中的两个角色：Server、Clients，他们需要完成以下的事情： Server：

1. 把最新的全局模型（Global Model）发送给一部分 Clients

2. 接收 Clients 训练后提交的模型更新（Model Updates），并进行聚合

3. 把聚合后的模型更新应用到全局模型上，得到新一轮的全局模型 Clients：

1. 从 Server 端获取最新的全局模型并进行训练（Local Training）

2. 发送训练后的模型更新给 Server

其中，聚合的方式最常用的是平均算法，算法如图4.2 所示：

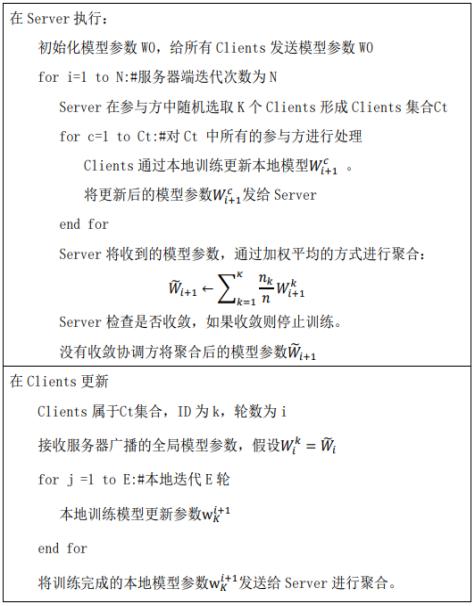


图 4.2 联邦学习平均算法的流程图

42

中国知网 https:Www  cnknet

**4.1.1 Mnist 数据集**

MNIST 数据集源自美国国家标准与技术研究所（NIST）。其训练集包括 250 个不同个体的手写数字，其中一半来自高中生，另一半来自美国人口普查局的 职员。测试集同样由手写数字组成，比例相同，但确保编写者与训练集不重叠。

这个数据集共包含七万张手写数字图像，6/7 的图像用于训练，1/7 的图像 用于测试。每张图像都是 28×28 像素的手写数字，范围从 0 到 9。数据集内容 具体如图 4.3 所示：



图 4.3 Mnist 数据集

**4.1.2 数据分发问题**

联邦学习的目的是保护用户的隐私数据，只需要在本地进行训练。McMahan 提出的 FedAvg 算法，算法的想法是不需要所有的客户端都上传局部模型去进行 聚合，而是从客户端中随机抽取一部分客户端接收上传的局部模型。这种抽取 部分客户端的方法会导致训练出来的模型与真实情况下的模型有一点差异，这 种差异受两个方面的影响。一方面是随机选取的部分客户端的数量，数量过少 不能很好的对模型进行优化。另一方面客户端上的数据的集中和分散性，过于 集中在某个特征上也会导致聚合后的模型出现问题。

在实际的联邦学习中， 存在两种需要考虑的问题： 一个是独立同分布 （Non-IID）问题，也就是客户端的隐私数据过于集中，导致每个客户端都有自

43

中国知网 https:Www  cnknet

己的偏好特征，这对模型的更新是不利的。另一个是选择的部分客户端占比的 问题，只有选取合适的比例才能使模型的更新与实际情况是相一致的，过多会 消耗很多不必要的时间，而过少又不能完整的体现模型的更新。

假设在这种情况下：有N个客户端参与联邦学习，他们所拥有的隐私数据 是不同的，那么联邦学习问题可以概述为：

 ............... 

wk 是第k个客户端的权重，每个客户端的权重都是要保证在(0,1)这个范围 的，同时所有客户端的权重相加是 1。假设第k个客户端拥有nk 条训练数据 : {xk, 1, xk,2, … , xk,nk}。本地目标函数Fk (w)的定义如下:

 ............... 

f(w; xk,j ) 是损失函数，它根据模型参数w和客户端的客户端拥有训练数据 : {xk, 1, xk,2, … , xk,nk}中的第w个，得出的xk,j 。可以在给定具体的模型参数w时，

检验预测结果和真实情况。

FedAvg 算法的主要思路是将经过聚合的新的全局模型同时分发给所有客户 端，客户端在接收到之后立刻进行本地训练获得局部模型。那么对于在第t轮的 联邦学习训练过程中，客户端将会进行下面的操作：

w+1 = w − ηt ∇Fk (w, ξ) ...............（4.3）

ηt 是学习率，代表联邦学习的学习效率。ξ是第t轮从客户端拥有训练数据 : {xk, 1, xk,2, … , xk,nk}中抽取到的一条数据，将这些数据收集起来可以得到新模型 

wt+E = ∑=1 wk 

FedAvg 算法虽然把模型分发给所有的客户端，但是只会从所有的客户端中

随机选择一部分客户端接收局部模型去进行模型的聚合{w+E, … , E }，对于选

取的部分客户端st 是由m个客户端构成的。如果选取的恰当，那么部分客户端 的模型更新和将所有客户端都进行更新是一致的，说明这种情况下客户端的选 取是无偏的：

=1 wkw ........ 

44

中国知网 https:Www  cnknet

接下来将使用 FedAvg 算法进行联邦学习的优化。

**4.1.3 实验设计**

使用的数据集：使用 Mnist 数据集中的 60000 个训练集样本去进行联邦学 习。首先生成 100 个 Clients，分别使用 IID（独立同分布）方式和 Non-IID （非独立同分布）方式为每个 Clients 分配 600 个样本。其中，客户端分心次 数，学习率等参数设置如表 4.1 所示。

表 4.1 参数设置

|  |  |  |
| --- | --- | --- |
| 参数 | 含义 | 参数设置 |
| g | Gpu | 0 |
| nc | 客户端的数量 | 100 |
| cf | 随机挑选的客户端的数量 | 10 |
| E | 客户端更新次数 | 1 |
| B | batchsize 大小 | 100 |
| lr | 学习率 | 0.01 |
| ncomm | 通信次数 | 200 |

1.使用 IID 分发数据：

IID（独立同分布）：在概率统计理论中的随机过程中，任何时刻的取值都 为随机变量，若这些随机变量服从同一分布且是相互独立的，则称这些随机变 量为独立同分布。

由于 MNIST 是图像分类的数据集都是 10 分类的数据，对于 IID 的方式，构 建每个客户端都包含上述数据集的 10 个分类，并且均匀分配，不聚集。

2.使用 Non-IID 分发数据：

Non-IID（非独立同分布）：在概率统计理论中，除了 IID 其余都是 Non- IID。

由于 MNIST 是图像分类的数据集都是 10 分类的数据，对于 Non-IID 的方式， 按照两种方式进行构建

1.Non-IID（1），构建每个客户端都包含 10 个分类数据中的其中一个分类 的全部

2.Non-IID（2），针对每个客户端都包含 10 个分类数据中的其中两个（每

45

中国知网 https:Www  cnknet

个数据集采样 50%）

对于 MNIST 数据集，B=100，E=1，其中 B 为 BatchSize（同时训练个数） 大小，E 为迭代次数。在两种不同方式下进行联邦学习的准确率如图 4.4：

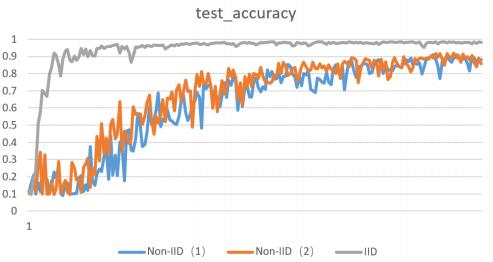


图 4.4 不同方式下进行联邦学习的准确率 根据上图我们可以得出：

在 IID 模式下的 B=100，E=1 时，从这个图中可以看出这种模式下准确率是 最高的，效果是最好的。

在 Non-IID（1）模式下的 B=100，E=1 时，从这个图中可以看出会比 IID 低很多.

在 Non-IID（2）模式下的 B=100，E=1 时，从这个图中可以看出这种模式 下比 IID 模型下性能差，但是会比 Non-IID（1）性能高。

所以可以得出结论在使用 Mnist 数据集时，使用 IID 模式去进行分发数据 得到的准确率是最好的。

其中，评价联邦学习的性能指标定义如下：

Accuracy =  ................ 

在这其中，TP (True Positive)和TN (True Negative)代表预测正确的样 本数量，而FP (False Positive)和FN (False Negative)表示预测错误的样本 数量。详细来说，TP是指被正确预测为正例的样本数，TN是指被正确预测为负 例的样本数，FP是指被错误地预测为正例的样本数，而FN是指被错误地预测为 负例的样本数。

46

中国知网 https:Www  cnknet

**4.1.4 Adam 梯度下降**

在联邦学习中通常使用加权平均的方式进行聚合，接下来使用加权平均和 梯度下降（Adam）的方式去进行实验，验证在不同情况下的准确率结果。 使用 Mnist 数据集中的 60000 个训练集样本去进行联邦学习，同时并生成 100 个 Clients。分别使用 IID（独立同分布）方式进行 200 轮通信，分别以加权平均 和梯度下降（Adam）的方式进行聚合，得到的联邦学习准确率如图4.5 所示：

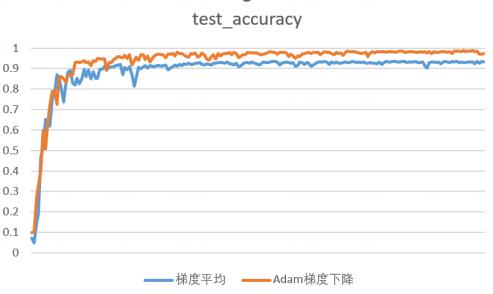


图 4.5 加权平均和 Adam 梯度下降模式下的准确率

从上图看出，使用 Adam 梯度下降方式与加权平均相比，准确率有明显提升， 说明使用这种方法对于联邦学习的准确率有积极作用

**4. 2 实验仿真平台搭建**

实验仿真平台的搭建如表 4.2 所示。在此基础上完成模型的搭建、训练以 及测试任务。

表 4.2 实验仿真平台配置

|  |  |
| --- | --- |
| W | 参数配置 |
| CPU | Ryzen9 5900HX |
| GPU | RTX 3080 |
| 开发语言 | Python 3.10 |
| 集成开发软件 | Spyder |

47

中国知网 https:Www  cnknet

**4. 3 利用数字水印增强隐私计算的功能性测试**

使用 Mnist 数据集中的 60000 个训练集样本去进行联邦学习。首先生成 100 个 Clients，分别使用 IID（独立同分布）方式（其中 B=100，E=1）为每个 Clients 分配 600 个样本，每轮选取 10 个 Clients 进行联邦学习，同时使用 Adam 梯度下降方式进行 200 论通信。分别在联邦学习中使用数字水印和不使用 数字水印方法。实验流程如下图 4.6 所示：

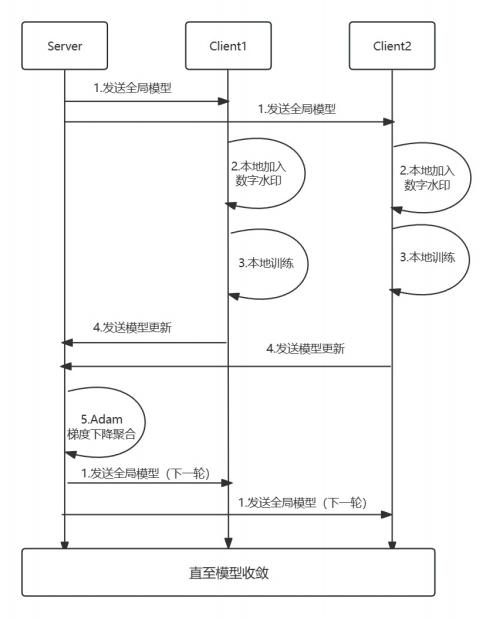


图 4.6 利用数字水印增强联邦学习的流程图

使用数字水印：在联邦学习开始，对本地图片使用第 3 章提出的数字水印 算法进行水印嵌入的操作，之后进行联邦学习。

48

中国知网 https:Www  cnknet

获得联邦学习的时间和准确率如表 4.3 和图 4.7 所示。

表 4.3 加入数字水印的联邦学习的性能

|  |  |  |
| --- | --- | --- |
|  | 未加数字水印 | 加数字水印 |
| 后 100 轮准确率均值 | 0.9789 | 0.9730 |
| 时间 | 71.0005 | 98.4699 |

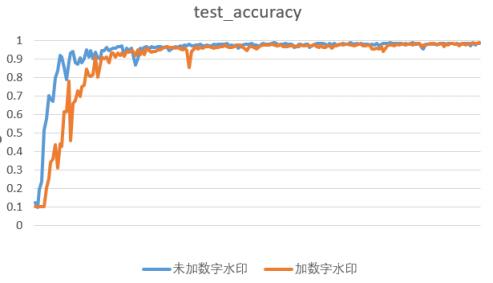


图 4.7 联邦学习的准确率

根据实验数据可得，在联邦学习中使用数字水印技术，不会影响联邦学习 的准确率，在 100 轮之后都可以保持在 97%以上的准确率，同时对比两种情况 下的实验时间，发现在加入数字水印后，会增加联邦学习的时间，但是会极大 的增强联邦学习的安全性，提高联邦学习的抗保密性攻击的能力，说明在联邦 学习中加入数字水印技术是可行的。

**4. 4 本章小结**

通过调整使用 Non-IID 的数据分发方式和 Adam 梯度下降聚合方式提升联 邦学习的准确率。最后在改进后的联邦学习中，使用数字水印技术，通过仿真 不影响联邦学习的准确性和时效性，同时提升了联邦学习的安全性，说明基于 数字水印的联邦学习方法是可行的。

49

中国知网 https:Www  cnknet

50

中国知网 https:Www  cnknet

**第** **5 章** **利用数字水印增强隐私计算的抗攻击测试**

在第 4 章验证了利用数字水印增强隐私计算安全性的可行性后，接下来在 第 5 章中将验证其安全性，特别是其抵御保密性攻击的能力。由于保密性攻击 试图窃取本地训练图片，因此第五章将从以下两个方面进行验证：

1.图片所有权验证：验证图片遭受噪声攻击、缩放攻击和裁剪攻击等后的 所有权。

2.抗攻击性能比较：与现有的几种水印算法比较，验证其抗攻击性能。

**5.1 实验仿真平台搭建**

实验仿真平台的搭建如表 5.1 所示。其中，CPU 型号为 Ryzen9 5900HX， GPU 型号为 RTX 3080，开发语言为 MATLAB 脚本语言，集成开发软件为 MATLAB R2023a。在此基础上完成模型的搭建、训练以及测试任务。

表 5.1 实验仿真平台配置

|  |  |
| --- | --- |
| 环境配置 | 参数配置 |
| CPU GPU  开发语言  集成开发软件 | Ryzen9 5900HX  RTX 3080  MATLAB 脚本语言  MATLAB R2023a |

**5.2 抗攻击测试结果与分析**

对利用数字水印增强联邦学习安全性方法进行抗攻击测试，分别进行了 JPEG 压缩攻击、噪声攻击、滤波攻击、缩放攻击、裁剪攻击和旋转攻击。

**5.2.1 JPEG 压缩攻击**

在数字水印的众多攻击方式中，其对 JPEG 压缩攻击的防御能力极为关键， 可以说是最重要的。这是因为在图像保存方面，大多数图像当前都采用 JPG 或 JPEG 格式；在图像传输方面，图像在平台传输时会经历平台自带的压缩算法， 这些算法在底层多少会使用到 JPEG 压缩。

51

中国知网 https:Www  cnknet

为了验证利用数字水印增强联邦学习安全性方法的抵抗 JPEG 压缩攻击的 能力，将已经嵌入水印的图像分别模拟 10%、40%和 70%的 PEG 压缩攻击，并观 测受到攻击后的图片与原始图片的相似度。同时对受攻击图片进行对应的提取 水印操作，并把提取得到的水印与原始水印进行对比，判断两者的相似程度， 同时使用 NC 去进行客观的评价。

接下来的实验都以 200×200 大小的 Lena 为载体图像和 100×100 大小的 “隐私计算 数字水印 ”二值图像为水印图像作为实验对象。

实验仿真的结果如表 5.2、图 5.1、图 5.2 和图 5.3 和所示：

表 5.2 JPEG 压缩后提取水印图像与原水印图像 NC 值

|  |  |  |  |
| --- | --- | --- | --- |
| 数字水印算法 | JPEG-10 | JPEG-40 | JPEG-70 |
| LL-LH | 0.9953 | 0.9997 | 0.9999 |
| LL-HL | 0.9957 | 0.9996 | 0.9999 |
| LL-HH | 0.9953 | 0.9997 | 0.9999 |



图 5.1 遭受 JPEG-10 攻击的图像和提取水印图像



图 5.2 遭受 JPEG-40 攻击的图像和提取水印图像 52

中国知网 https:Www  cnknet



图 5.3 遭受 JPEG-70 攻击的图像和提取水印图像

从图 5.1、图 5.2 和图 5.3 中可以观察到，使用这种数字水印的图片即使 在遭受到 JPEG 压缩率为 10%、40%还是 70%攻击之后，即使图片发生了变化，但 是都可以根据提取出来的数字水印进行所有权认证。由于其具有很好的鲁棒性， 在受攻击之后，提取出来的水印图像与原始的水印图像的 NC 值都超过了 0.995， 表明这种方法有很强的抵抗 JPEG 压缩攻击的能力。

综合仿真结果来看，基于 DWT-SVD 的双重鲁棒数字水印算法在抵御 JPEG 压 缩攻击方面表现出色，从而证明了利用数字水印增强隐私计算安全性的方法在 对抗 JPEG 压缩攻击方面具有优异的性能。

**5.2.2 噪声攻击**

为了验证利用数字水印增强联邦学习安全性方法的抵抗噪声攻击的能力， 将已经嵌入水印的图像分别模拟 0.001 方差的高斯噪声、0.001 方差的椒盐噪 声以及 0.005 方差的散斑噪声攻击，并观测受到攻击后的图片与原始图片的相 似度。同时对受攻击图片进行对应的提取水印操作，并把提取得到的水印与原 始水印进行对比，判断两者的相似程度，同时使用 NC 去进行客观的评价。

实验仿真的结果如表 5.3、图 5.4、图 5.5 和图 5.6 所示：

表 5.3 噪声攻击后提取水印图像与原水印图像 NC 值

|  |  |  |  |
| --- | --- | --- | --- |
| 数字水印算法 | 高斯噪声攻击 | 椒盐噪声攻击 | 散斑噪声攻击 |
| LL-LH | 0.9924 | 0.9987 | 0.9864 |
| LL-HL | 0.9926 | 0.9992 | 0.9859 |
| LL-HH | 0.9927 | 0.9991 | 0.9869 |

53

中国知网 https:Www  cnknet



图 5.4 遭受 0.001 方差的高斯噪声攻击的图像和提取水印图像



图 5.5 遭受 0.001 方差的椒盐噪声攻击的图像和提取水印图像



图 5.6 遭受 0.005 方差的散斑噪声攻击的图像和提取水印图像

54

中国知网 https:Www  cnknet

观察图 5.4、图 5.5 和图 5.6 发现，使用这种数字水印的图片即使在遭受 到高斯噪声、椒盐噪声以及散斑噪声攻击之后，即使图片发生了变化，但是都 可以根据提取出来的数字水印进行所有权认证。由于其具有很好的鲁棒性，在 受攻击之后，提取出来的水印图像与原始的水印图像的 NC 值都超过了 0.98， 表明这种方法有很强的抵抗 JPEG 压缩攻击的能力。

综合仿真结果，基于 DWT-SVD 的双重鲁棒数字水印算法在抵抗噪声攻击 （包括高斯噪声、椒盐噪声和散斑噪声）方面表现出良好的能力，从而证明了 利用数字水印增强隐私计算安全性的方法在对抗噪声攻击方面具有优异的性能。

**5.2.3 滤波攻击**

为了验证利用数字水印增强联邦学习安全性方法的抵抗滤波攻击的能力，

将已经嵌入水印的图像分别模拟 3×3 的高斯低通滤波、3×3 的均值滤波和 3× 3 的中值滤波攻击，并观测受到攻击后的图片与原始图片的相似度。同时对受 攻击图片进行对应的提取水印操作，并把提取得到的水印与原始水印进行对比， 判断两者的相似程度，同时使用 NC 去进行客观的评价。

实验仿真的结果如表 5.4、图 5.7、图 5.8 和图 5.9 和所示：

表 5.4 滤波攻击后提取水印图像与原水印图像 NC 值

|  |  |  |  |
| --- | --- | --- | --- |
| 数字水印算法 | 高斯低通滤波 | 均值滤波 | 中值滤波 |
| LL-LH | 0.9838 | 0.9732 | 0.9875 |
| LL-HL | 0.9838 | 0.9733 | 0.9874 |
| LL-HH | 0.9836 | 0.9731 | 0.9875 |



图 5.7 遭受 3×3 高斯低通滤波攻击的图像和提取水印图像

55

中国知网 https:Www  cnknet

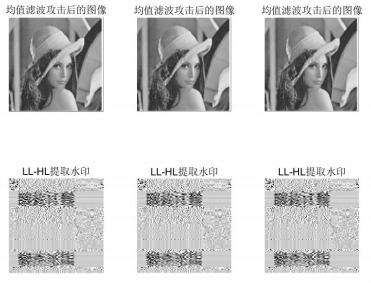


图 5.8 遭受 3×3 均值滤波攻击的图像和提取水印图像



图 5.9 遭受3×3 中值滤波攻击的图像和提取水印图像

观察图 5.7、图 5.8 和图 5.9 发现，使用这种数字水印的图片即使在遭受 到高斯低通滤波、均值滤波和中值滤波攻击之后，即使图片发生了变化，但是 都可以根据提取出来的数字水印进行所有权认证。由于其具有很好的鲁棒性，

在受攻击之后，提取出来的水印图像与原始的水印图像的 NC 值都超过了 0.973， 表明这种方法有很强的抵抗滤波攻击的能力。

综合仿真结果，基于 DWT-SVD 的双重鲁棒数字水印算法在抵抗滤波攻击 （包括高斯低通滤波、均值滤波和中值滤波攻击）方面展现出良好的性能，从 而证明了利用数字水印增强隐私计算安全性的方法在抵抗缩放攻击方面同样具 有优异的性能。

56

中国知网 https:Www  cnknet

**5.2.4 缩放攻击**

为了验证利用数字水印增强联邦学习安全性方法的抵抗缩放攻击的能力， 将已经嵌入水印的图像分别模拟 0.5 倍和 1.5 倍的缩放攻击，并观测受到攻击 后的图片与原始图片的相似度。同时对受攻击图片进行对应的提取水印操作， 并把提取得到的水印与原始水印进行对比，判断两者的相似程度，同时使用 NC 去进行客观的评价。实验仿真的结果如表 5.4、图 5.10 和图 5.11 和所示：

表 5.4 缩放攻击后提取水印图像与原水印图像 NC 值

0.5 倍缩放攻击 · 1.5 倍缩放攻击 LL-LH 0.9996 0.9996

LL-HL 0.9996 0.9996 LL-HH 0.9996 0.9996

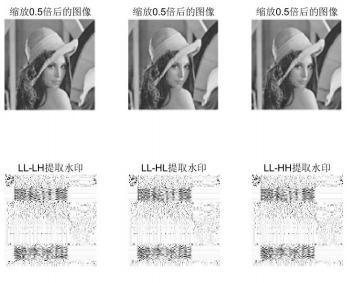


图 5.10 遭受 0.5 倍缩放攻击的图像和提取水印图像



图 5.11 遭受 1.5 倍缩放攻击的图像和提取水印图像

57

中国知网 https:Www  cnknet

观察图 5.10 和图 5.11 发现，使用这种数字水印的图片即使在遭受到 0.5 倍和 1.5 倍的缩放攻击之后，即使图片发生了变化，但是都可以根据提取出来 的数字水印进行所有权认证。由于其具有很好的鲁棒性，在受攻击之后，提取 出来的水印图像与原始的水印图像的 NC 值都超过了 0.9996，表明这种方法有 很强的抵抗 JPEG 压缩攻击的能力。

综合仿真结果表明，基于 DWT-SVD 的双重鲁棒数字水印算法在抵抗缩放攻 击（包括 0.5 倍缩放和 1.5 倍缩放）方面展现出良好的性能，从而证明了利用 数字水印增强隐私计算安全性的方法在抵抗缩放攻击方面同样具有优异的性能。

**5.2.5 裁剪攻击**

为了验证利用数字水印增强联邦学习安全性方法的抵抗裁剪攻击的能力，

将已经嵌入水印的图像分别模拟中心裁剪攻击和边缘裁剪攻击，并观测受到攻 击后的图片与原始图片的相似度。同时对受攻击图片进行对应的提取水印操作， 并把提取得到的水印与原始水印进行对比，判断两者的相似程度，同时使用 NC 去进行客观的评价。

实验仿真的结果如表 5.5、图 5.12 和图 5.13 和所示：

表 5.5 裁剪攻击后提取水印图像与原水印图像 NC 值

中心裁剪攻击 · 边缘裁剪攻击 LL-LH 0.9638 0.9986

LL-HL 0.9639 0.9986 LL-HH 0.9635 0.9986

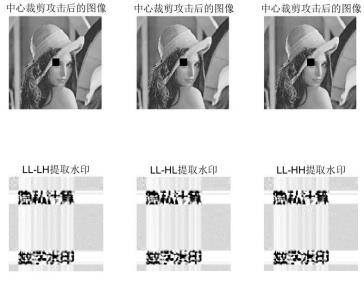


图 5.12 遭受中心裁剪攻击攻击的图像和提取水印图像

58

中国知网 https:Www  cnknet



图 5.13 遭受边缘裁剪攻击攻击的图像和提取水印图像

观察图 5.12 和图 5.13 发现，使用这种数字水印的图片即使在遭受到中心 裁剪攻击和边缘裁剪攻击之后，即使图片发生了变化，但是都可以根据提取出 来的数字水印进行所有权认证。由于其具有很好的鲁棒性，在受攻击之后，提 取出来的水印图像与原始的水印图像的 NC 值都超过了 0.96，表明这种方法有 很强的抵抗 JPEG 压缩攻击的能力。

综合仿真结果表明，基于 DWT-SVD 的双重鲁棒数字水印算法在抵抗裁剪攻 击（包括中心裁剪和边缘裁剪攻击）方面表现良好，说明利用数字水印增强隐 私计算安全性的方法在抵抗旋转攻击方面同样具有优秀的性能。

**5.2.6 旋转攻击**

为了验证利用数字水印增强联邦学习安全性方法的抵抗旋转攻击的能力，

将已经嵌入水印的图像分别模拟 10°、20°和 30°的旋转攻击，并观测受到攻 击后的图片与原始图片的相似度。同时对受攻击图片进行对应的提取水印操作， 并把提取得到的水印与原始水印进行对比，判断两者的相似程度，同时使用 NC 去进行客观的评价。

实验仿真的结果如表 5.6、图 5.14、图 5.15 和图 5.16 和所示： 表 5.6 旋转攻击后提取水印图像与原水印图像 NC 值

|  |  |  |  |
| --- | --- | --- | --- |
| 数字水印算法 | 10°旋转攻击 | 20°旋转攻击 | 30°旋转攻击 |
| LL-LH | 0.7969 | 0.5880 | 0.5058 |
| LL-HL | 0.7972 | 0.5823 | 0.5060 |
| LL-HH | 0.7699 | 0.5880 | 0.5057 |

59

中国知网 https:Www  cnknet

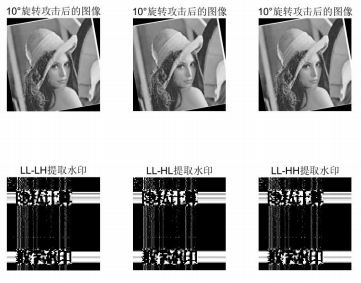


图 5.14 遭受旋转攻击-10°攻击的图像和提取水印图像



图 5.15 遭受旋转攻击-20°攻击的图像和提取水印图像

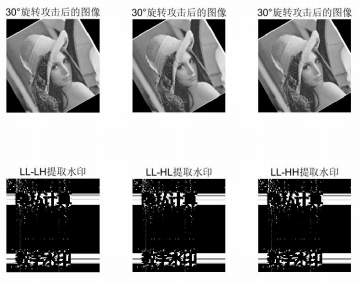


图 5.16 遭受旋转攻击-30°攻击的图像和提取水印图像

60

中国知网 https:Www  cnknet

观察图 5.14、图 5.15 和图 5.16 发现，使用这种数字水印的图片即使在遭 受到旋转攻击之后，即使图片发生了变化，但是都可以根据提取出来的数字水 印进行所有权认证。由于其具有很好的鲁棒性，在受攻击之后，提取出来的水 印图像与原始的水印图像的 NC 值都超过了 0.5，表明这种方法有一定的抵抗 JPEG 压缩攻击的能力。

综合仿真结果表明，基于 DWT-SVD 的双重鲁棒数字水印算法在抵抗旋转攻 击（包括 10°、20°和 30°旋转攻击）方面表现出一定的能力，从而证明了利 用数字水印增强隐私计算安全性的方法在抵抗旋转攻击方面具有一定的效果。

**5.3 抗攻击性能对比分析**

为了进一步验证利用数字水印提升隐私计算安全性的效果，特别是在抵抗 攻击方面的性能，即数字水印算法的鲁棒性，本文将提出的数字水印算法与一 些经典的数字水印算法进行了对比分析。采用归一化相关系数（NC）作为定量 指标来评估各个算法的性能，相关结果见表 5.7。

表 5.7 本文算法与现有算法提取水印的 NC 值对比

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Liu[63] | Thanki[64] | Huang[65] | LL-LH | LL-HL | LL-HH |
| JEPG-30 | 0.9463 | 0.9495 | 0.9321 | **0.9994** | **0.9994** | **0.9994** |
| JEPG-60 | 0.9655 | 0.9741 | 0.9702 | **0.9998** | **0.9998** | **0.9998** |
| JEPG-90 | 0.9899 | 0.9910 | 0.9911 | **1.0000** | **1.0000** | **1.0000** |
| 高斯噪声 | 0.8227 | 0.9500 | 0.9035 | **0.9924** | **0.9926** | **0.9927** |
| 椒盐噪声 | 0.8860 | 0.9843 | 0.8299 | **0.9987** | **0.9992** | **0.9992** |
| 高斯低通滤波攻击 | 0.9143 | 0.9916 | 0.9522 | **0.9838** | **0.9838** | **0.9836** |
| 均值滤波攻击 | 0.8651 | 0.2258 | 0.8003 | **0.9732** | **0.9733** | **0.9731** |
| 中值滤波攻击 | 0.8916 | 0.4576 | 0.9258 | **0.9875** | **0.9874** | **0.9875** |
| 0.5 倍缩放 | 0.6452 | 0.5051 | 0.9532 | **0.9996** | **0.9996** | **0.9996** |
| 中心裁剪攻击 | 0.9365 | **0.9795** | 0.9547 | 0.9638 | 0.9639 | 0.9635 |
| 30 °旋转攻击 | 0.5322 | 0.4798 | **0.6543** | 0.5058 | 0.5060 | 0.5057 |

和现有的数字水印算法进行对比，可以发现本文提出的方法在应对各种攻 击时的抗攻击能力更高，可以保证即使受到攻击也可以凭借提取的数字水印进 行所有权的验证。因此，本算法在抵抗攻击方面表现出色，从而证明了利用数 字水印增强隐私计算安全性的有效性。

61

中国知网 https:Www  cnknet

**5.4 本章小结**

本章在第四章的基础上，对利用数字水印可以增强联邦学习的安全性进行 了抗攻击测试。对嵌入水印的图像分别进行了 JPEG 压缩攻击、噪声攻击、滤波 攻击、缩放攻击、裁剪攻击和旋转攻击。同时，本文将该方法与现有几种 （Liu[63] 、Thanki[64]和 Huang[65]提出的）水印算法进行了比较，通过实验，可以 看出数字水印在遭受攻击之后，提取出的水印图像效果比现有的几种方法高。 当联邦学习面临保密性攻击导致数据泄露时，即便攻击者对图片进行了修改， 仍然可以从修改后的图片中恢复出嵌入的水印，实现追踪溯源功能，从而提升 联邦学习的安全性。

因此，基于数字水印的联邦学习方法不仅在不影响联邦学习准确率的前提 下提高了安全性和保密性，还具有一定的抵御攻击能力。

62

中国知网 https:Www  cnknet

**第** **6 章** **总结与展望**

**6.1 全文总结**

本文通过利用数字水印可以增强联邦学习的安全性，着手在不影响联邦学 习的准确性和时效性的前提下，利用数字水印提高联邦学习的安全性和保密性 同时使联邦学习具有抗攻击能力，具体工作总结如下：

（1）提出一种基于 DWT（离散小波变换）和 SVD（奇异值分解）的双重 鲁棒数字水印算法。首先将图像进行离散小波变换，提高嵌入水印图像的不可 见性，之后利用 Arnold 变换对水印图像进行加密，提高了水印图像的安全性， 同时通过使用奇异值分解进一步增强水印图像的鲁棒性。仿真结果表明，使用 该算法与基于 SVD 的数字水印算法和基于 DWT-SVD 的单鲁棒数字水印算法相 比，性能有明显的提升。

（2）提高联邦学习的效率，验证该方法的可行性。在联邦学习中通过使用 Non-IID 的数据分发方式和 Adam 梯度下降聚合方式提升联邦学习的准确率。通 过实验表明，在调整后的联邦学习中使用数字水印方法，不影响联邦学习的准 确性和时效性，说明利用数字水印增强联邦学习安全性是可行的。

（3）对利用数字水印增强隐私计算联邦学习安全性的方法的抗攻击测试， 分别进行了 JPEG 压缩攻击、噪声攻击、滤波攻击、缩放攻击、裁剪攻击、旋 转攻击。在受到攻击的情况下提取出来的水印图像的效果都比较好。因此当联 邦学习遇到保密性攻击导致泄露数据时，即使攻击方对图片进行操作，但从修 改后的图片中仍然可以恢复出嵌入的水印，从而实现追踪溯源的功能，提升联 邦学习的安全性。该方法不仅在不影响联邦学习的准确性和时效性的前提下提 高联邦学习的安全性和保密性，而且使联邦学习具有抗攻击能力。

**6.2 未来展望**

（1）在未来的研究过程中，计划对数字水印技术进一步研究，尝试融合其 他数字水印技术去提升抗攻击能力，同时减少算法消耗的时间。以保证在应用

63

中国知网 https:Www  cnknet

在联邦学习中时，可以得到更高的准确率和低时耗性。

（2）本文只针对隐私计算中的联邦学习进行了数字水印的应用，但是隐私 计算的其他实现方式如多方安全计算、可信执行环境和多方中介计算并没有进 行验证，未来可以从这三方面进行实验验证。

（3）在利用数字水印增强联邦学习的安全性研究中，主要通过对数字水印 的不可见性以及联邦学习的框架进行优化来提升准确性，没有通过对数字水印 的嵌入过程和联邦学习的训练时间提升联邦学习训练的时效性，因此未来可以 着力于改进数字水印嵌入算法和优化联邦学习框架。

64

中国知网 https:Www  cnknet

**参考文献**

[1] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-lev el performance on imagenet classification[C]//Proceedings of the IEEE internati onal conference on computer vision. 2015: 1026-1034.

[2] Bao H, Dong L, Wei F. Beit: Bert pre-training of image transformers[J]. arXiv p

reprint arXiv:2106.08254, 2021.

[3] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommend ations[C]//Proceedings of the 10th ACM conference on recommender systems. 2 016: 191-198.

[4] Naumov M, Mudigere D, Shi H J M, et al. Deep learning recommendation mod el for personalization and recommendation systems[J]. arXiv preprint arXiv:190 6.00091, 2019.

[5] Gomez-Uribe C A, Hunt N. The netflix recommender system: Algorithms, busin ess value, and innovation[J]. ACM Transactions on Management Information Sy stems (TMIS), 2015, 6(4): 1-19.

[6] Hatcher W G, Yu W. A survey of deep learning: Platforms, applications and eme

rging research trends[J]. IEEE Access, 2018, 6: 24411-24432.

[7] Deng J, Dong W, SocherR, et al. Imagenet: A large-scale hierarchical image dat abase[C]//2009 IEEE conference on computer vision and pattern recognition. Ie ee, 2009: 248-255.

[8] 杨瑞仙，李兴芳，王栋，臧国全. 隐私计算的溯源、现状及展望[J].情报理 论与实践,2023,第 46 卷(7): 158-167

[9] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of d

eep networks from decentralized data[C]//Artificial intelligence and statistics. P MLR, 2017: 1273-1282.

[10]FREDRIKSON M, JHA S, RISTENPART T. Model inversion attacks that exploi t confidence information and basic countermea sures[C]//Proceedings of the 22n

65

中国知网 https:Www  cnknet

d ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2015: 1322-1333.

[11]TOLPEGIN V, TRUEX S, GURSOY M E, et al. Data poisoning attacks against federated learning systems[C]//European Symposium on Research in Computer Security. Berlin: Springer, 2020: 480-501. .

[12]ZHANG J L, CHEN J J, WU D, et al. Poisoning attack in federated learning usi ng generative adversarial nets[C]//Proceedings of 2019 18th IEEE International Conference on Trust, Security and Privacy In Computing and Communications/ 13th IEEE International Conference on Big Data Science and Engineering (Trus tCom/BigDataSE). Pisca taway: IEEE Press, 2019: 374-380.

[13]MOTHUKURI V, PARIZI R M, POURIYEH S, et al. A survey on security and privacy of federated learning[J]. Future Generation Computer Systems, 2021, 1 15: 619-640. .

[14]PROUDFOOT D. Anthropomorphism and AI: Turing’s much misunderstood im itation game[J]. Artificial Intelligence, 2011, 175(5-6): 950-957.

[15]ZHANG J L, CHEN B, CHENG X, et al. PoisonGAN: generative poisoning atta cks against federated learning in edge computing systems[J]. IEEE Internet of T hings Journal, 2021, 8(5): 3310-3322.

[16]BAGDASARYAN E, VEIT A, HUA Y, et al. How to backdoor fede rated learni ng[C]//International Conference on Artificial Intelligence and Statistics. New Yo rk: PMLR, 2020: 2938-2948.

[17]XU G W, LI H W, LIU S, et al. VerifyNet: secure and verifiable federated learni ng[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 911- 926.

[18]LU Y L, HUANG X H, DAI YY, et al. Blockchain and federated learning for pr ivacy-preserved data sharing in industrial IoT[J]. IEEE Transactions on Industri al Informatics, 2020, 16(6): 4177-4186.

[19]YANG Q, LIU Y, CHEN T J, et al. Federated machine learning[J]. ACM Transa ctions on Intelligent Systems and Technology, 2019, 10(2): 1-19.

66

中国知网 https:Www  cnknet

[20]Li Y,Ouyang K,LiN,et al.A blockchain-assisted intelligent transportation system

promoting data services with privacy protection[J].Sensors,2020,20(9):2483.

[21]赵阔,邢永恒. 区块链技术驱动下的物联网安全研究综述[J].信息网络安全,20 17(05):1-6.、

[22]Wu H T,Tsai C W.Toward blockchains for health-care systems:Applying the bili nearpairing technology to ensure privacy protection and accuracy in data sharin g[J].IEEE Consumer Electronics Magazine,2018,7(4):65-71.

[23]Li Y,Zhang G,Zhu J,et al.Privacy Protection Model for Blockchain Data Sharing Basedon zk-SNARK[C]//International Conference of Pioneering Computer Sci entists,Engineers and Educators.Springer,Singapore,2021:229-239.

[24]Kim H,Park J,Bennis M,et al.Blockchained on-device federated learning[J].IEE ECommunications Letters,2019,24(6):1279-1283.

[25]Y,Gao L,Luan T H,et al.Decentralized privacy using blockchain-enabled federat edlearning in fog computing[J].IEEE Internet of Things Journal,2020,7(6):5171- 5183.

[26]Pokhrel S R,Choi J.Federated learning with blockchain for autonomous vehicles: Analysis and design challenges[J].IEEE Transactions on Communications,2020, 68(8):4734-4746.

[27]Zhao Y,Zhao J,Jiang L,et al.Privacy-preserving blockchain-based federated learn

ing forIoT devices[J].IEEE Internet of Things Journal,2020,8(3):1817-1829.

[28]Suyi Li,Yong Cheng,Yang Liu,et al.Abnormal Client Behavior Detection in Fed eratedLearning[C]//International Workshop on Federated Learning for User Priv acy and Data Confidentiality in Conjunction with NeurIPS 2019(FL-NeurIPS'1 9).Vancouver Convention Center,Vancouver,BC,Canada.

[29]Ramanan P,Nakayama K.Baffle:Blockchain based aggregator free federated lear ning[C]//2020 IEEE International Conference on Blockchain(Blockchain).IEEE, 2020:72-81.

67

中国知网 https:Www  cnknet

[30]Lu Y,Huang X,Dai Y,et al.Blockchain and federated learning for privacy-preserv ed data sharing in industrial IoT[J].IEEE Transactions on Industrial Informatics, 2019,16(6):4177-4186.

[31]Lu Y,Huang X,Zhang K,et al.Blockchain empowered asynchronous federated le arning for secure data sharing in internet of vehicles[J].IEEE Transactions on Ve hicular Technology,2020,69(4):4298-4311.

[32]B. G. A. Tekgul et al, "WAFFLE: Watermarking in federated learning," in 2021, . DOI: 10.1109/SRDS53918.2021.00038.

[33]B. Han et al, "Application of Robust Zero-Watermarking Scheme Based on Fede rated Learning for Securing the Healthcare Data," IEEE Journal of Biomedical a nd Health Informatics, vol. 27, (2),pp. 804-813, 2023.

[34]W. Yang et al, "Watermarking in Secure Federated Learning: A Verification Fra mework Based on Client-Side Backdooring," ACM Transactions on Intelligent S ystems and Technology, vol. 15, (1),pp. 1-25, 2023.

[35]Tirkel A Z, Rankin G A, Schyndel R V, et al. Electronic Wartermark[C]. Macqua rie Univercity: Digital Image Computing, Technology and Applications(DICTA 93),1993. 666-673.

[36]Mansouri A, Mahmoudi-Aznavehb A. Toward a secure video watermarking in c ompressed domain[J]. Journal of Information Security and Applications, 2019, 4 8: 102370.

[37]Sun Y F, Wang J Y, Huang H G, et al. Research on scalable video watermarking

algorithm based on H.264 compressed domain[J]. Optik, 2021, 227:165911.

[38]Singh R, AshokA, raswat M. Optimised robust watermarking technique using C KGSA in DCT -SVD domain[J]. IET Image Processing, 2020,14(10): 2052-20 63.

[39]Li D, Che X, Luo W et al. Digital watermarking scheme for colour remote sensi ng image based on quaternion wavelet transform and tensor decomposition[J]. Mathematical Methods in the Applied Sciences,2019,42(4): 4664-4678.

68

中国知网 https:Www  cnknet

[40]Chen L X, Bai W T, Yao Z Q. A Secure and Privacy ‐Preserving Watermark Ba sed Medical Image Sharing Method[J]. Chinese Journal of Electronics,2020, 29 (5): 819-825.

[41]Sathya S PA, Ramakrishnan S. Non-redundant frame identification and keyfram e selection in DWT-PCA domain for authentication of video[J]. IET Image Proc essing, 2020,14(2): 366-375.

[42]Al-JUMEIL Y ，ALLOGHANI M ，ALANI M M.A systematicreview on the st atus and progress of homomorphic encryption technologies［J］.

[43]YANG Qiang ，LIU Yang ，CHEN Tianjian ，et al.Federatedmachine learning:c oncept and applications［J］．ACM Transactions on Intelligent Systems and Te chnology(TIST) ，2019 ， 10(2):1-19.

[44]邵航，高思琪，钟离，等. 同态加密在隐私计算中的应用综述［J］．信息通 信技术与政策，2022 ，48(8):75-88.

[45]孙爽，李晓会，刘妍，等.不同场景的联邦学习安全与隐私保护研究综述 ［J］．计算机应用研究，2021 ，38(12):3527-3534.

[46]宁振宇，张锋巍，施巍松.基于边缘计算的可信执行环境研究［J］．计算机 研究与发展，2019 ，56(7):1441-

[47]N. Wang et al, "A blockchain based privacy-preserving federated learning sche

me for Internet of Vehicles," Digital Communications and Networks, vol. 10, (1), pp. 126-134, 2024.

[48]JEONG E, OH S, KIM H, et al. Communication-efficient on-device machine lea rning: federated distillation and augmentation under non-iid private data[J]. arXi vpreprint, 2018, arXiv: 1811.11479.

[49]Ionescu R T, Popescu M. Knowledge transfer between computer vision and text mining[J]. Advances in Computer Vision and Pattern Recognition. Springer Inte rnational Publishing, 2016.

[50]李伟君．基于图像认证的脆弱与半脆弱水印技术研究[D] ．大连：大连理 工大学，2007 .

69

中国知网 https:Www  cnknet

[51]A1-Gindy,Ahmed N,Ahmaa ，et a1 ．A new blind image watermarking of hand written

[52]张宪海，杨永田.一种双重变换域图像半脆弱水印算法[J].计算机工程与应 用,2007,(9): 59

[53]Cox IJ, Kilian J, Leighton T, Shamoon T. Secure spread spectrum watermarking

for images, Audio and Video. In Proceedings of ICIP, Lausane, Switzerland, 19 96, No.3, 243-246.

[54]Piva A. DCT-based watermark recovering without resorting to the uncorrupted o riginal images. In Proceedings of ICIP, 1997, No.1, 520-523.

[55]Hsu C.T, Wu j. J Hidden signature in images. In Proceedings of ICIP, 1996, No. 3, 223-226.

[56]W. Zeng and B. Liu, "A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images," IEEE Trans actions on Image Processing, vol. 8, (11),pp. 1534-1548, 1999.

[57]Wang H J M, Su P C, Kuo C C J. Wavelet-based digital image watermarking. O ptics Express, 1998, 3(12):491-496.吴晓平,祁亨年.未知信号传播速度的TDO A测量远近场统一定位方法[J].传感技术学报,2020,33(07):1049-1055.

[58]Kunder D, Hatzinakos D. Digital watermarking using multiresolution wavelet d

ecomposion. In Proceedings of IEEE International Conference on Acoustics, Sp eech and Signal Processing, Seattle, Washington, 1998, No.5, 2969-2972.

[59]Chen B., Wornell G.W.. Quantization index modulation: a class of provably goo d methods for digital watermarking and information embedding[J]. IEEE Transa ctions on Information Theory, 2001, 47(4): 1423-1443.

[60]Cox I.J., Kilian J., Leighton F.T., et al.. Secure spread spectrum watermarking fo r multimedia[J]. IEEE Transactions on Image Processing,1997, 6(12): P.1673-16 87.

[61]刘西林. 数字图像变换域鲁棒性水印算法研究[D]. 东南大学博士学位论文, 2017.

70

中国知网 https:Www  cnknet

[62] A. Anand and A. K. Singh, "An improved DWT-SVD domain watermarking for medical information security," Computer Communications, vol. 152,pp. 72-80, 2020.

[63]冀峰. 基于几何不变性的鲁棒图像水印算法[D]. 西安电子科技大学博士学 位论文，2012.

[64]Van R., Tirkel A.Z., Osborne C.F.. A digital watermark[C]. Proceedings of 1st In ternational Conference on Image Processing, Austin TX, USA, 1994.

[65]Liu Y, Tang S, Liu R, et al. Secure and robust digital image watermarking schem e using logistic and RSA encryption[J]. Expert Systems with Applications, 2018, 97: 95-105.

[66]Ko H J, Huang C T, Horng G, et al. Robust and blind image watermarking in D CT domain using inter-block coefficient correlation[J]. Information Sciences, 20 20, 517: 128-147.

71

中国知网 https:Www  cnknet