软件学报 ISSN 1000-9825, CODEN RUXUEW
2025,36(7):3087−3108 [doi: 10.13328/j.cnki.jos.007339] [CSTR: 32375.14.jos.007339]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

# 结合特征生成与重放的可扩展安全虹膜识别*

赵冬冬[1], 宋宝刚[1], 廖虎成[2], 闫 江[1], 向剑文[1]

[1](武汉理工大学 计算机与人工智能学院, 湖北 武汉 430070)

[2](云南省农村信用社科技结算中心, 云南 昆明 650228)

通信作者: 向剑文, E-mail: jwxiang@whut.edu.cn

**摘 要**: 随着信息技术的快速发展, 安全认证技术成为个人隐私和数据安全的重要保障. 其中, 虹膜识别技术凭借其出色的准确性和稳定性, 被广泛应用于系统访问控制、医疗保健以及司法实践等领域. 然而用户的虹膜特征数据泄露, 就是永久性丢失, 无法进行更改或者撤销. 因此, 虹膜特征数据的隐私保护尤为重要. 随着神经网络技术在图像处理上体现的突出性能, 基于神经网络的安全虹膜识别方案被提出, 在保护隐私数据的同时保持了识别系统的高性能. 然而, 面对不断变化的数据和环境, 安全虹膜识别方案需要具备有效的可扩展性, 即识别方案应当能够在新的用户注册下依旧保持性能. 但大多数现有基于神经网络的安全虹膜识别方案研究并未考虑方案的可扩展性. 针对上述问题, 提出了基于生成特征重放的安全增量虹膜识别 (generative feature replay-based secure incremental iris recognition, GFR-SIR) 方法和基于隐私保护模板重放的安全增量虹膜识别 (privacy-preserving template replay-based secure incremental iris recognition, PTR-SIR) 方法. 具体而言, GFR-SIR 方法通过生成特征重放和特征蒸馏技术, 缓解神经网络扩展过程中对以往任务知识的遗忘, 并采用改进的 TNCB 方法来保护虹膜特征数据的隐私. PTR-SIR 方法保存了以往任务中通过 TNCB 方法转换得到的隐私保护模板, 并在当前任务的模型训练中重放这些模板, 以实现识别方案的可扩展性. 实验结果表明, 在完成 5 轮扩展任务后, GFR-SIR 和 PTR-SIR 在 CASIA-Iris-Lamp 数据集上的识别准确率分别达到了 68.32% 和 98.49%, 比微调方法分别提升了 58.49% 和 88.66%. 分析表明, GFR-SIR 方法由于未保存以往任务的数据, 在安全性和模型训练效率方面具有明显优势; PTR-SIR 方法则在维持识别性能方面更为出色, 但其安全性和效率低于 GFR-SIR.

**关键词**: 隐私保护; 虹膜识别; 转换网络; 增量学习

**中图法分类号**: TP18

中文引用格式: 赵冬冬, 宋宝刚, 廖虎成, 闫江, 向剑文. 结合特征生成与重放的可扩展安全虹膜识别. 软件学报, 2025, 36(7): 3087–3108. http://www.jos.org.cn/1000-9825/7339.htm

英文引用格式: Zhao DD, Song BG, Liao HC, Yan J, Xiang JW. Scalable Secure Iris Recognition Combining Feature Generation and Replay. Ruan Jian Xue Bao/Journal of Software, 2025, 36(7): 3087–3108 (in Chinese). http://www.jos.org.cn/1000-9825/7339.htm

## Scalable Secure Iris Recognition Combining Feature Generation and Replay

ZHAO Dong-Dong[1], SONG Bao-Gang[1], LIAO Hu-Cheng[2], YAN Jiang[1], XIANG Jian-Wen[1]

[1](School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China)

[2](Yunnan Rural Credit Cooperatives Technology Settlement Center, Kunming 650228, China)

**Abstract**: With the rapid development of information technology, security authentication technology has become a crucial safeguard for personal privacy and data security. Among them, iris recognition technology, with its outstanding accuracy and stability, is widely applied

* 基金项目: 国家重点研发计划 (2022YFC3321102); 湖北省重点研发计划 (2022BAA050)
本文由“新兴软件与系统的可信赖性与安全”专题特约编辑向剑文教授、陈厅教授、杨珉教授、周俊伟教授推荐.
收稿时间: 2024-08-26; 修改时间: 2024-10-15; 采用时间: 2024-11-25; jos 在线出版时间: 2024-12-10
CNKI 网络首发时间: 2025-04-09

in fields such as system access control, healthcare, and judicial practices. However, once the iris feature data of a user is leaked, it means permanent loss, as it cannot be changed or revoked. Therefore, the privacy protection of iris feature data is particularly important. With the prominent performance of neural network technology in image processing, secure iris recognition schemes based on neural networks have been proposed, which maintain the high performance of recognition systems while protecting privacy data. However, in the face of constantly changing data and environments, secure iris recognition schemes are required to have effective scalability, that is, the recognition scheme should be able to maintain performance with new user registrations. However, most of the existing research on neural network-based secure iris recognition schemes does not consider the scalability of the schemes. Aiming at the above problems, the generative feature replay-based secure incremental iris recognition (GFR-SIR) method and the privacy-preserving template replay-based secure incremental iris recognition (PTR-SIR) method are proposed in this study. Specifically, the GFR-SIR method uses generative feature replay and feature distillation techniques to alleviate the forgetting of previous task knowledge during the expansion of neural networks and adopts the improved TNCB method to protect the privacy of iris feature data. The PTR-SIR method preserves the privacy-protecting templates obtained through the TNCB method in previous tasks and replays these templates during the model training of the current task to achieve the scalability of the recognition scheme. Experimental results show that after completing 5 rounds of expansion tasks, the recognition accuracy of GFR-SIR and PTR-SIR on the CASIA-Iris-Lamp dataset reaches 68.32% and 98.49% respectively, which is an improvement of 58.49% and 88.66% compared with the fine-tuning method. The analysis indicates that the GFR-SIR method has significant advantages in terms of security and model training efficiency since the data of previous tasks is not saved; while the PTR-SIR method is more outstanding in maintaining recognition performance, but its security and efficiency are lower than those of GFR-SIR.

**Key words**: privacy protection; iris recognition; transformation network; incremental learning

随着计算机网络和通信技术的飞速发展, 基于人脸、指纹和虹膜等生物特征的身份认证技术日益成熟. 其中, 虹膜识别技术得益于虹膜纹理特征的高可区分性和长期稳定性, 成为一种高准确率和高可靠性的生物特征识别方法. 该技术凭借这些优势在公共安全、民生服务以及司法应用等关键领域得到了广泛应用[1]. 例如, 阿拉伯联合酋长国自 2001 年以来一直使用虹膜识别技术来进行边境管制[2]. 在新冠肺炎疫情期间, 穿戴医疗防护服的医护人员由于面部遮挡无法使用人脸识别核验身份, 因此, 在方舱医院出入口通过虹膜识别技术进行非接触式身份识别[1]. 然而, 虹膜识别在给人们的生活带来诸多便利的同时, 也带来了隐私数据泄露的问题. 例如, 印度曾有超过 8 亿用户的隐私数据遭到泄露, 其中包括用户的虹膜和指纹等生物特征数据. 隐私数据的泄露严重威胁到了用户的个人隐私、心理健康以及财产安全, 引起了公众对安全生物特征识别技术的高度关注.

虹膜在人出生后的几个月内形成并定型, 在一生中几乎保持不变. 一旦用户的虹膜特征数据泄露, 就是永久性丢失, 无法进行更改或者撤销. 因此, 有效的安全虹膜识别策略对于保护用户隐私至关重要. 国际标准 ISO/IEC[3]中对于安全虹膜识别方法的安全性要求进行了进一步的规定. 具体来说: (1) 不可逆性, 要求很难从相应的受保护生物特征模板重建原始生物特征; (2) 不可链接性, 要求由相同生物特征生成的不同受保护生物特征模板应该不可区分; (3) 可撤销性, 要求当受保护生物特征模板泄露时, 应该很容易生成新的受保护生物特征模板.

为了满足上述 3 种性质, 许多针对生物特征数据的隐私保护方案被提出, 这些方案可以被分为: (1) 可撤销生物特征识别 (cancellable biometrics, CB); (2) 生物特征密码系统 (biometric cryptosystem, BCS). 可撤销生物特征识别技术对原始生物特征数据进行不可逆的转换, 采用不同参数对同一生物特征获得不同受保护的生物特征模板, 并通过比较受保护的生物特征模板之间的相似度与设定阈值的方式来进行匹配. 然而, 许多可撤销的生物识别方法已被证明容易受到各种攻击[4–7], 这表明这些方法无法满足不可逆性和不可链接性的要求. 在生物特征密码系统中, 原始生物特征数据与加密密钥结合形成辅助数据, 并存储在数据库中. 用户结合辅助数据恢复加密密钥, 并验证密钥的有效性间接实现匹配. 然而 BCS 引入纠错码 (error correcting code, ECC) 来克服生物识别中的类内差异, 并实现允许某些错误的模糊识别. 在许多工作中已经表明[8–11], 纠错码的引入以及辅助数据的存储, 会对 BCS 造成严重的隐私问题. 此外, 现有的 CB 和 BCS 这些传统安全生物识别方案在一些存在挑战的数据集上无法得到理想的性能.

随着神经网络在图像处理上体现的突出性能, 基于神经网络的安全生物识别方案得到了发展. 基于神经网络的安全生物识别方案相比传统的安全生物识别方案取得了更好的性能. 然而神经网络的使用也为这些方法带来了一个不可避免的问题: 随着用户数据和应用场景的不断更新, 这些方案的扩展性面临了巨大的挑战. 具体来说, 训

练好的神经网络仅仅适用于出现在训练集中的用户, 无法扩展到未出现在训练集中的用户. 一种简单看似合理的方案是存储先前任务的训练数据, 并在新任务出现时, 结合新数据重新训练模型. 然而这样的方式需要原始生物数据的存储, 从而会导致原始生物信息的泄露. 基于以上原因, 在不对神经网络训练数据进行隐私保护的情况下, 当新用户注册时, 通常只能基于新的训练数据和前一任务的模型完成当前任务的模型训练. 这种情况下得到的当前任务模型在以往任务上的识别准确率会显著下降, 严重影响模型的正常使用. 因此, 设计一种基于神经网络的可扩展安全生物识别方案十分重要.

针对上述问题, 本文提出了基于生成特征重放的安全可扩展虹膜识别 (generative feature replay-based secure incremental iris recognition, GFR-SIR) 方法和基于隐私保护模板重放的安全可扩展虹膜识别 (privacy-preserving template replay-based secure incremental iris recognition, PTR-SIR) 方法. 这两种方案都对神经网络的训练数据采用 TNCB 进行处理保护隐私数据实现神经网络的扩展. 具体而言, GFR-SIR 方法通过生成特征重放和特征蒸馏技术, 缓解模型扩展过程中对以往任务知识的遗忘, 并采用改进的 TNCB 方法来保护虹膜特征数据的隐私. PTR-SIR 方法通过保存以往任务中通过 TNCB 方法转换得到的隐私保护模板, 并在新任务的模型训练中重放这些模板, 以实现神经网络的可扩展性并保证高性能.

本文的贡献可以总结为以下 3 点.

(1) 针对存储空间有限的应用场景, 本文提出了基于生成特征重放的安全可扩展虹膜识别方法, 在不保存训练数据的情况下, 实现了高安全性的可扩展虹膜识别方案.

(2) 针对识别性能要求较高的应用场景, 本文提出了基于隐私保护模板重放的安全可扩展虹膜识别方法, 在仅保存受保护训练数据的情况下, 实现了高性能的安全可扩展虹膜识别方案.

(3) 本文在 CASIA-Iris-Lamp 和 CASIA-Iris-Thousand 数据集上进行大量的实验来验证本文所提出的两种方案的性能. 实验结果表明所提出的方案在保护隐私数据的前提下得到了较好的性能.

## 1 相关工作

针对生物特征数据的隐私保护, 已有的工作主要分为两大类: 生物特征密码系统[12]和可撤销生物特征识别[13].

生物特征密码系统[12]通常通过将生物特征数据和密钥进行安全的绑定或者从生物特征数据中生成密钥的方式来完成隐私保护. 主要方案包括模糊金库[14]、模糊承诺[15]和模糊提取器[16]. 具体来说, 模糊保险库方案使用一组无序点来保护密钥, 创建一个保险库需要多项式重建技术来检索密钥. 然而, 许多研究[17−19]表明, 模糊金库方案可能会泄露原始生物特征数据, 并且容易受到交叉匹配攻击. 模糊承诺方案将 ECC 与密码学技术相结合, 通过将生物特征数据与随机码字相关联来创建承诺. 然后使用纠错技术恢复码字以进行验证. 然而, 一些研究[8,9]中已经证明, 模糊承诺方案容易受到可解码攻击. 随后, Kelkboom 等人[20]提出一种改进的方案来防止交叉匹配. 这之后, Rathgeb 等人[10]提出模糊承诺方案进行统计攻击, 成功检索了存储的密钥, 说明模糊承诺方案仍然存在隐私泄露的风险. 模糊提取器方案从生物特征数据中提取一致的随机字符串和辅助数据. 在重建过程中, 将辅助数据与相似的生物特征数据相结合, 重建出一致的随机字符串. 该方案使用提取的随机字符串作为加密密钥. 然而, Blanton 等人[11]指出, 当多个草图暴露时, 模糊提取方案难以确保不可链接性和不可逆性, 从而导致潜在的隐私泄露.

可撤销生物特征识别[13]通过可重复的不可逆变换方法将原始生物特征数据转换为隐私保护模板来保证原始生物特征数据的安全. Rathgeb 等人[21]提出了一种将二进制虹膜编码映射到布隆过滤器实现免对齐的可撤销虹膜识别方法, 该方法将二进制虹膜编码矩阵分别按照水平和竖直方向划分为不同的块和字进行模板映射. 然而, Hermans 等人[4]对基于布隆过滤器的安全虹膜识别方法的安全性进行了评估, 并指出该方法在不可链接性方面存在安全性问题. Jin 等人[22]提出了一种基于局部敏感哈希的可撤销生物识别方法, 称为 IoM 哈希. 该方法通过将实值的生物特征向量映射为排序之后的最大索引值来生成隐私保护模板. 但是, Ghammam 等人[5]对于基于 IoM 哈希的方法进行了攻击, 并指出该方法无法抵抗链接攻击和认证攻击. 此外, Sadhya 等人[23]也基于局部敏感哈希的理论框架提出了一种安全虹膜识别方法, 通过随机位采样生成类内紧凑的隐私保护模板, 该方法实现了较高的识别性能.

Lai 等人[24]根据最小哈希可用于进行相似性检索的原理, 提出了一种称为 IFO 哈希的可撤销虹膜模板生成方法, 该方法还通过 Hadamard 积和取模运算进一步提升安全性. 但是, Dong 等人[6]基于遗传算法对 IFO 方法发起了原像攻击以构建隐私保护模板对应原始生物特征的近似值, 文中指出 IFO 方法无法抵御该原像攻击, 并且该方法也可能会泄露视觉信息. Zhao 等人[25]基于负数据库这一信息安全技术提出了一种安全虹膜识别方法, 该方法将原始生物特征数据转化为负数据保存为隐私保护模板, 通过负数据库恢复原始数据这一过程的 NP 难解来保证虹膜特征模板的安全. 但 Ouda 等人[7]对负虹膜识别方法的安全性进行了评估, 指出该方法容易遭受数据逆向攻击、链接攻击和多重记录攻击.

由于传统的 CB 和 BCS 方案在一些充满挑战的数据集上 (如 CASIA-Iris-Thousand、CASIA-Iris-Lamp) 无法达到令人满意的性能, 近 10 年来也有很多基于神经网络的安全虹膜识别方法被提出. Abdellatef 等人[26]提出了一种基于神经网络的可撤销生物特征识别方法, 该方法实现了不错的识别性能, 但是文中并未对安全性进行详细分析. Kim 等人[27]提出了一种基于深度学习的端到端多模态可撤销生物识别方案, 该方法将人脸和眼部区域特征融合之后通过哈希和压缩模块隐私保护模板. 但上述两个方法由于使用了用户特定的密钥, 存在由于用户特定密钥被盗而造成隐私泄露的风险, 也无法满足安全性要求. 由于安全性和高性能的需求, 一种基于神经网络与预定义二进制码的安全生物识别方法被提出, 在文献 [28] 中, 为每个用户分配一个随机的最大熵二进制码, 并训练 CNN 将用户的图像映射到该二进制码. 最后, 对二进制代码应用加密哈希函数生成最终的保护模板. 在文献 [29] 中, 以前使用的最大熵二进制码被低密度奇偶校验 (low-density parity-check, LDPC) 码所取代. 该方案利用 LDPC 码的纠错能力, 增强了对类内变化的鲁棒性. 这两种方案虽然通过加密哈希函数实现了较高的安全性, 并取得了较好的识别性能, 但在可扩展性上面临着重大挑战. 当新用户注册时, 需要重新训练整个模型, 这意味着先前的训练数据需要被存储, 然而这样会带来很大的隐私风险.

根据上述讨论, 可以看到结合神经网络的生物识别方案能够同时实现安全性与高性能, 然而这些方法由于神经网络的使用, 在可扩展性上存在问题. 因此, 本文提出了 GFR-SIR 和 PTR-SIR 两种方案来实现结合神经网络的生物识别方案的可扩展性.

## 2 预备知识

### 2.1 转换网络

在转换网络的训练过程中, 对于输入图像 $x_i$, 首先通过转换网络对其进行转换得到隐私保护的图像 $\tilde{x}_i$, 并通过特征重建损失[30]来度量图像转换前后的特征损失, 计算方法如下:

$$L_{\text{feat}}(\tilde{x}_i, x_i) = \frac{1}{C_j H_j W_j} \left\| \phi_j(\hat{x}_i) - \phi_j(x_i) \right\|_2^2 \tag{1}$$

其中, $\phi_j(x_i)$ 为一个尺寸为 $C_j \times H_j \times W_j$ 的特征图, 该特征图是将 $x_i$ 输入到卷积神经网络之后在第 $j$ 个卷积层的激活值. 接下来, 将经过转换之后的图像 $\tilde{x}_i$ 输入到训练好的识别网络中得到预测标签 $\hat{y}_i$, 并通过下列公式计算多分类交叉熵损失:

$$L_{\text{class}}(\hat{y}_i, y_i) = -\sum_{k=l}^{n} y_{i,k} \log \hat{y}_{i,k} \tag{2}$$

其中, $n$ 为类别数量, $y_{i,k}$ 和 $\hat{y}_{i,k}$ 分别表示第 $i$ 个样本的真实标签为 $k$ 和预测标签属于类别 $k$ 的概率. 将上述特征重建损失和分类损失进行组合可以得到特征转换的联合损失 $L_{\text{trans}}$, 计算方法如下:

$$L_{\text{trans}} = -\sum_{k=1}^{n} y_{i,k} \log \hat{y}_{i,k} - \alpha \cdot \left( \frac{l}{C_j H_j W_j} \left\| \phi_j(\hat{x}_i) - \phi_j(x_i) \right\|_2^2 \right) \tag{3}$$

其中, 变量 $\alpha$ 为一个权重参数, 用于控制特征重建损失所占的比例. 最终通过最小化联合损失 $L_{\text{trans}}$ 完成模型的训练, 具体方法可以总结为:

$$\min\frac{1}{m}\sum_{i=1}^{m}L_{\text{trans}}(x_i,h_\theta(x_i),y_i) \tag{4}$$

Zhao 等人[31]将转换网络的方法引入到虹膜识别领域, 提出了一种基于转换网络的安全虹膜识别方法 TNCB, 该方法通过分块置换、取模和正反合并[32]运算对安全性进行了改进, 方法架构如图 1 所示.

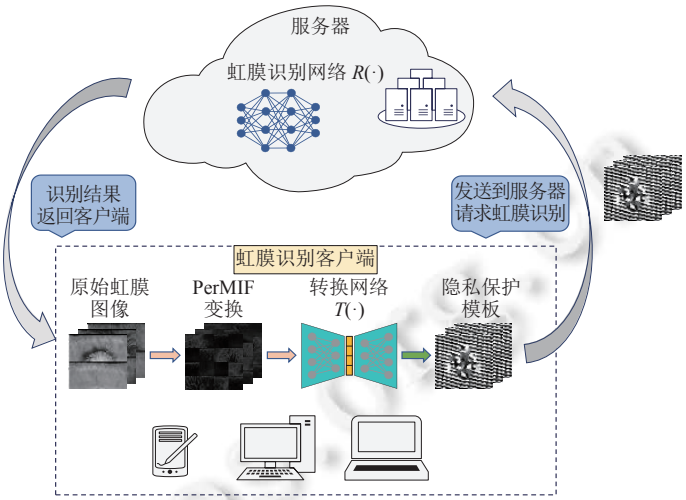


图 1 基于转换网络的安全虹膜识别方法架构

在 TNCB 方法第 1 阶段的变换过程中, 先根据应用特定的置换参数对原始虹膜图像进行了分块置换, 并进行逐像素的取模和正反合并运算, 具体的 PerMIF 算法流程如算法 1 所示.

**算法 1**. PerMIF 算法.

输入: 虹膜图像 $I$、块大小 $b$、取模阈值 $\tau$、正反合并运算阈值 $\gamma$、应用的置换参数 $p$、虹膜图像宽度 $W$ 和高度 $H$;
输出: 隐私保护图像 $X$.

1. **for** $i=1$ to $H$ **do**
2. **for** $j=1$ to $W$ **do** // 根据置换参数 $p$ 对输入图像 $I$ 进行分块置换得到 $P$
3. 计算 $P_{i,j}$
4. $M_{i,j}=P_{i,j}\bmod\tau$ // 取模运算
5. **if** $M_{i,j}\leqslant\gamma$ **do** // 正反合并运算
6. $X_{i,j}=M_{i,j}$
7. **else**
8. $X_{i,j}=\tau-1-M_{i,j}$
9. **end if**
10. **end for**
11. **end for**
12. **return** $X$

在第 2 阶段, 先将上述运算得到的图像用于虹膜识别网络的训练, 接下来根据识别网络通过对抗的思想训练转换网络, 转换网络的目的是在保持识别性能的同时防止攻击者重构出第 1 阶段变换之后的图像. 在完成识别网络和转换网络的训练之后, 将转换网络部署到客户端, 用于将经过第 1 阶段变换的虹膜图像转换为最终的隐私保护模板. 而将识别网络部署到服务器, 用于对客户端发送过来的隐私保护模板进行识别. 最终, 服务器将识别结果

发送回客户端.

### 2.2 生成特征重放

由于在复杂的任务中, 基于图像生成的方法往往难以生成高质量的图像, 因此 Liu 等人[33]提出了基于生成特征重放的类别增量学习方法, 通过重放以往任务的特征以缓解当前任务训练过程中模型对于以往任务知识的遗忘. 具体来说, 该方法中训练了一个特征生成器 $G_t$ 将特征的条件分布 $p_u=(u|c)$ 建模为 $\hat{u}=G_t(c,z)$, 在训练新任务时, 从该分布中进行采样得到生成的特征与新数据提取的特征一起完成分类器的训练. 该工作中采用 WGAN 进行特征建模和生成重放, 以实现增量学习. 其中, 生成器和鉴别器在训练过程中的损失函数如下:

$$L_{G_t}^{\mathrm{WGAN}}(X_t)=-E_{z\sim p_z,\ c\in C_t}[D_t(c,G_t(c,z))] \tag{5}$$

$$L_{D_t}^{\mathrm{WGAN}}(X_t)=+E_{z\sim p_z,\ c\in C_t}[D_t(c,G_t(c,z))]-E_{u\sim D_t}[D_t(c,F_t(x))] \tag{6}$$

在训练过程中还加入了重放对齐损失, 当以一个给定的先前类别 $c$ 和一个潜在向量 $z$ 为条件时, 该重放损失鼓励当前任务的生成器 $G_t$ 重放与以往任务生成器 $G_{t-1}$ 一致的特征. 重放对齐损失的计算方法如下:

$$L_{G_t}^{\mathrm{RA}}=\sum_{j=1}^{t-1}\sum_{c\in C_j}E_{z\sim p_z}[\|G_t(c,z)-G_{t-1}(c,z)\|_2^2] \tag{7}$$

在模型对抗训练过程中, 交替更新鉴别器和生成器, 其中, 鉴别器的训练通过最小化判别损失 $L_{D_t}^{\mathrm{WGAN}}(X_t)$ 来完成, 而生成器的训练通过同时最小化 $L_{G_t}^{\mathrm{WGAN}}(X_t)$ 和 $L_{G_t}^{\mathrm{RA}}$ 的联合损失完成.

此外, 为了缓解在新任务训练过程中特征提取器的知识遗忘, 该工作中还通过以往任务的特征提取器 $F_{t-1}$ 对当前任务的特征提取器 $F_t$ 进行了特征蒸馏, 该特征蒸馏损失的计算方法为:

$$L_{F_t}^{\mathrm{FD}}(X_t)=E_{x\sim X_t}[\|F_t(x)-F_{t-1}(x)\|_2] \tag{8}$$

## 3 本文方法

### 3.1 基于生成特征重放的安全增量虹膜识别

(1) 方法整体架构. 虹膜识别的实际应用场景中往往有注册新用户的需求, 这种需求类似增量学习任务, 在本文中将这种需求定义为扩展任务. 而在基于神经网络的安全虹膜识别模型扩展的过程中往往存在灾难性遗忘问题, 进而引起扩展任务的识别准确率显著下降. 对此, 本文引入生成特征重放技术对 TNCB 方法进行改进, 提出了基于生成特征重放的安全可扩展虹膜识别方法 (GFR-SIR), 架构如图 2 所示.

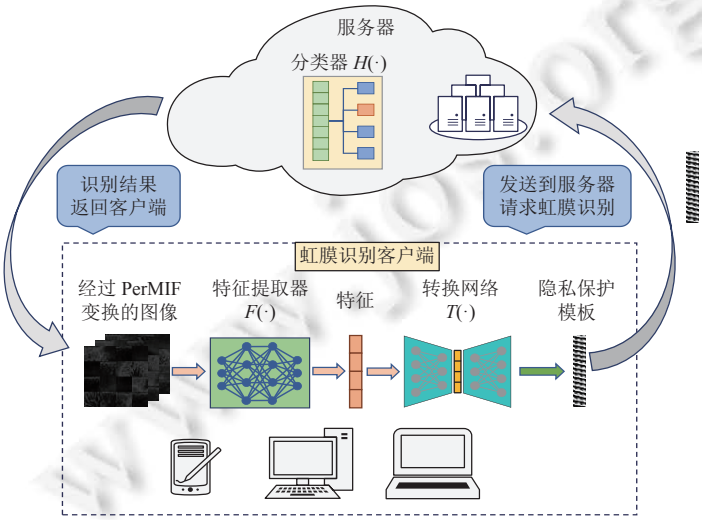


图 2 基于生成特征重放的安全增量虹膜识别方法模型部署架构

在注册阶段, 将原始虹膜图像经过 PerMIF 算法处理, 并在得到的图像上完成虹膜识别网络以及特征转换网络的训练模型训练完成之后, 与 TNCB 方法不同的是, GFR-SIR 将识别网络分为特征提取器和分类器两个部分进行分开部署. 其中, 将特征提取器和特征转换网络都部署到客户端. 在验证阶段, 对于用户提交用于请求识别的虹膜图像, 首先对其进行 PerMIF 算法处理, 并将变换得到的图像输入到特征提取器中得到特征向量. 随后通过转换网络对特征进行转换得到隐私保护模板, 并将该隐私保护模板发送到服务器请求识别. 服务器端则直接将分类器模型部署到服务器端, 通过分类器对接收到的隐私保护模板进行分类识别.

(2) 基于生成特征重放的扩展方法. 本方法的模型扩展训练算法如算法 2 所示. 该算法以不同任务经过初步隐私保护处理的数据集作为输入, 完成模型的扩展训练. 在初始任务中, 直接通过数据集训练得到对应的特征提取器、分类器以及特征生成器, 随后通过特征提取器和分类器进行特征转换网络的训练, 用于将特征转换为最终的隐私保护模板. 而对于扩展阶段的模型训练, 则是先通过生成器重放以往任务的特征, 并将重放特征和当前数据集中提取的特征一起完成当前任务的特征提取器、分类器和特征生成器的训练. 最后基于特征提取器提取的特征和重放的特征共同完成特征转换网络的扩展训练.

**算法 2**. GFR-SIR 算法.

输入: 当前任务 $t$ 的数据集 $D_t$ ($t \in \{0,1,\ldots,K\}$), 其中, $D_t=(X_t,C_t)$, $X_t$ 为经过 PerMIF 变换的隐私保护图像;
输出: 生成器 $G_t$、特征提取器 $F_t$、分类器 $H_t$ 和特征转换网络 $T_t$.

1. **if** $t=0$ **do**
2. 在初始任务数据集 $D_0$ 上训练特征提取器 $F_0$、分类器 $H_0$ 和特征生成器 $G_0$
3. 根据 $F_0$ 和 $H_0$ 训练特征转换网络 $T_0$ // $T_0$ 用于将特征向量转换为隐私保护模板
4. **else** // 通过以往任务的生成器生成特征, $t'$ 表示以往所有任务的编号
5. $\hat{u}_{t'}=G_{t-1}(C_{t'},z)$
6. 通过数据集 $D_t$ 和生成的特征 $\hat{u}_{t'}$ 训练 $F_t$、$H_t$ 和 $G_t$
7. 提取特征 $u_t=F_t(x_i)$, 其中, $x_i \in D_t$ // $T_t$ 用于将特征向量转换为隐私保护模板
8. 通过特征 $u_t$ 和 $\hat{u}_{t'}$ 以及分类器模型 $H_t$ 训练特征转换网络 $T_t$
9. **end if**
10. **return** $G_t$、$F_t$、$H_t$ 和 $T_t$

为了缓解安全虹膜识别模型扩展过程中由于知识遗忘而导致的性能下降, 本文对生成特征重放的类别增量方法进行了改进, 提出了基于生成特征重放的安全可扩展虹膜识别方法. 相对于原来的方法, 本文方法中对所有参与训练的数据进行了隐私保护, 确保生成特征重放的引入不会导致原始生物数据的泄露. 在识别网络扩展训练过程中涉及特征提取器、分类器和特征生成器这 3 个模块, 设 $t$ 为当前任务的编号, $F_t$ 和 $H_t$ 分别表示当前任务 $t$ 的特征提取器和分类器, 同理, $F_{t-1}$ 和 $G_{t-1}$ 分别为第 $t-1$ 个任务的特征提取器和特征生成器. 在当前任务的特征提取器 $F_t$ 和分类器 $H_t$ 训练过程中, 本方法通过在以往任务数据上训练的特征生成器重放以往任务的特征与当前任务提取的特征一起训练分类器, 从而缓解分类器的知识遗忘. 此外, 为了缓解特征提取器扩展训练过程中的知识遗忘, 通过特征蒸馏对其进行训练, 训练过程鼓励当前任务的特征提取器与以往任务的特征提取器提取到的特征保持一致. 在本方案中特征生成器采用 WGAN 进行特征建模和生成重放, 根据公式 (5) 和公式 (6) 分别计算生成器和判别器的损失, 考虑到随着扩展任务的进行, 生成器 $G_t$ 可能无法保证重放与以往任务生成器 $G_{t-1}$ 一致的特征, 训练过程中根据公式 (7) 加入了重放对齐损失, 鼓励当前任务的生成器 $G_t$ 重放与以往任务生成器 $G_{t-1}$ 一致的特征.

特征提取器与分类器具体的扩展训练过程如图 3 所示. 模型训练之前, 先将原始虹膜图像通过 PerMIF 算法进行第 1 阶段的变换得到模型的训练数据. 具体来说, 将原始虹膜图像经过分块置换、取模和正反合并运算, 再通过得到的隐私保护图像训练特征提取器和分类器. 在训练过程中, 将识别网络分为特征提取器 $F_t$ 和分类器 $H_t$ 两个部分, 先将经过正反合并运算之后的训练数据分别输入到当前任务 $t$ ($t \geqslant 1$) 和前一任务 $t-1$ 的特征提取器中得到 $u_t$

和 $u_{t-1}$, 并根据公式 (8) 计算得到特征蒸馏损失 $L_{F_t}^{\mathrm{FD}}$. 接下来通过生成器重放以往任务的特征 $\hat{u}_{t'}$, 并将其与特征 $u_t$ 一起输入到分类器中进行训练, 其中, $t'$ 表示以往所有任务的任务编号. 接下来通过公式 (2) 计算分类器输出的预测标签 $\hat{y}$ 和真实标签 $y$ 之间的分类损失 $L_{\mathrm{class}}$ 以及中心损失 $L_{\mathrm{cent}}$, 并根据分类损失、中心损失和特征蒸馏损失通过反向传播调整参数, 进而完成整个识别网络的增量训练. 中心损失的计算方法为:

$$L_{\mathrm{cent}}=\frac{1}{2}\sum_{i=1}^{m}\left\|x_i-c_{y_i}\right\|_2^2 \tag{9}$$

其中, $x_i$ 为第 $i$ 个图像, $c_{y_i}$ 表示类别 $y_i$ 的深度特征中心.

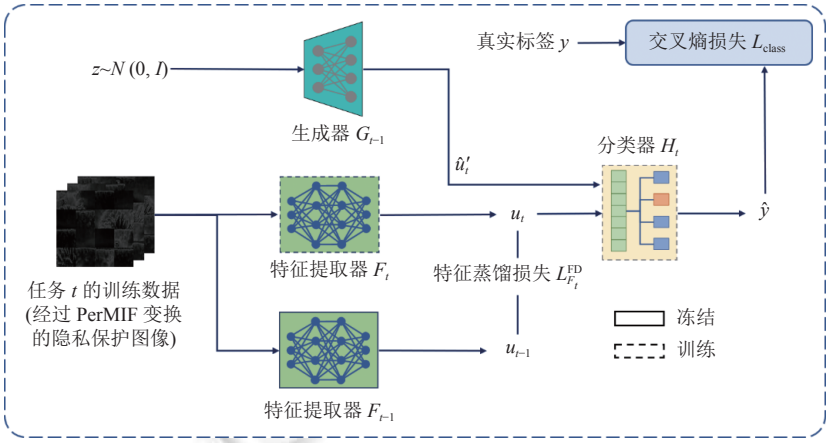


图 3 特征提取器和分类器的增量训练过程

在上一阶段完成特征提取器和分类器的训练之后, 使用当前任务训练好的特征提取器 $F_t$ 和分类器 $H_t$ 以及上一任务训练好的生成器 $H_{t-1}$ 一起训练转换网络 $T_t$, 模型训练过程如图 4 所示. 转换网络用于在客户端对特征提取器提取的特征进行不可逆变换, 进而生成最终的隐私保护模板.

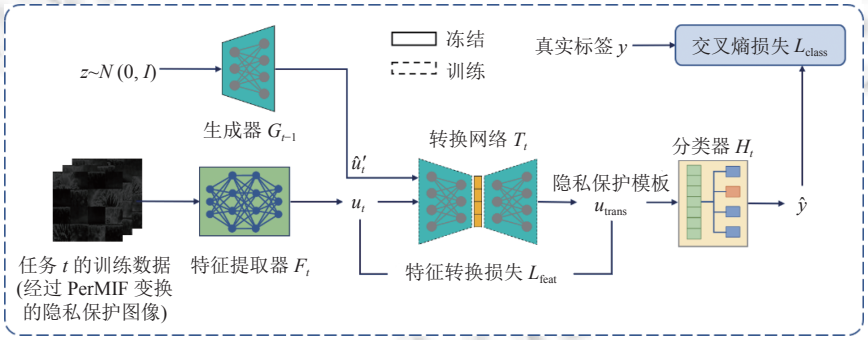


图 4 特征转换网络的增量训练过程

具体来说, 对于当前任务 $Task_t$ 的转换网络, 首先将 PerMIF 算法转换的隐私保护图像输入到特征提取网络中得到特征 $u_t$, 并将提取到的特征和生成器重放的特征 $\hat{u}_{t'}$ 一起输入到转换网络中进行转换得到特征 $u_{\mathrm{trans}}$. 转换前后特征之间的损失 $L_{\mathrm{feat}}$ 为均方误差. 随后将转换之后的特征 $u_{\mathrm{trans}}$ 输入到分类器中得到预测标签 $\hat{y}$, 并通过公式 (2) 计算预测标签与真实标签之间的损失 $L_{\mathrm{class}}$. 结合特征转换损失和分类损失可以得到转换网络的联合损失 $L_{\mathrm{trans}}$, 计算方法如下:

$$L_{\mathrm{trans}}=L_{\mathrm{class}}-\alpha\cdot L_{\mathrm{feat}} \tag{10}$$

其中, $\alpha$ 为权重参数, 用于在训练过程中控制特征转换损失所占的权重. 该联合损失表明转换网络的训练过程是最大化特征转换损失, 同时最小化分类器的分类损失, 从而达到既保持识别准确率, 又对特征进行隐私保护的目的.

### 3.2 基于隐私保护模板重放的安全增量虹膜识别

(1) 方法整体架构. 第 3.1 节介绍了本文提出的基于生成特征重放的安全可扩展虹膜识别方法, 尽管该方法在很大程度上降低了性能下降的幅度, 但对于一些对识别准确率要求很高的应用场景, 该方法还存在提升的空间. 保存先前任务的训练数据, 在当前任务训练时重放可以大幅提高模型的识别准确率. 然而考虑到安全生物识别场景下, 用户相关的未受保护数据无法进行保存, 这样的策略看似无法实现. 针对这个问题, 本文提出了一种简单而有效的方案基于隐私保护模板重放的安全增量虹膜识别方法, 称为 PTR-SIR, 使得保存先前任务的训练数据成为可能. 本方法在识别架构上与 TNCB 类似, 不同的地方在于, 为了能够保存先前任务的训练数据, 在新任务中仅对识别网络进行扩展训练, 而将特征转换模型按照任务划分为多个, 为每次扩展任务的用户单独训练一个转换模型并进行保存. 这样做的原因是经过转换网络处理得到的隐私图像是受保护的, 可以被存储起来用于后续的扩展训练 (即识别网络的训练数据是受保护的, 而转换网络的训练数据存在隐私泄露的风险). 在注册阶段, 将原始虹膜图像经过 PerMIF 算法处理, 并在得到的图像上完成特征转换网络的训练以及虹膜识别网络的扩展训练, 并将训练好的特征转换网络与隐私保护进行保存. 在验证阶段, 对于用户提交用于请求识别的虹膜图像, 首先对其进行 PerMIF 算法处理, 再将其输入到对应的特征转换网络进行转换得到隐私保护模板, 并将该隐私保护模板发送到识别网络请求识别. 值得注意的是, PTR-SIR 方法中使用的所有训练数据均经过 PerMIF 变换保护, 而并非原始训练数据, 这样做的目的是为了对用户相关的数据进行保护, 使得能够存储先前任务的训练数据而不泄露原始训练数据.

(2) 基于隐私保护模板重放的扩展方法. PTR-SIR 方法具体的模型扩展训练算法如算法 3 所示. 该算法同样以不同任务的数据集作为输入, 这些数据集为经过 PerMIF 算法变换的隐私保护图像. 对于初始任务模型, 首先在数据集上完成识别网络 $R_0$ 和图像转换网络 $T_0$ 的训练, 并通过转换网络将隐私保护图像转换为隐私保护模板后保存到服务器. 在后续任务模型的增量训练过程中, 首先训练一个转换网络, 将当前任务经过 PerMIF 变换的隐私保护图像转换为最终的隐私保护模板, 并与服务器中保存的之前所有任务的隐私保护模板共同完成识别网络 $R_t$ 的增量训练.

**算法 3**. PTR-SIR 算法.

输入: 当前任务 $t$ 的数据集 $D_t$ ($t \in \{0,1,\ldots,K\}$), 其中, $D_t=(X_t,C_t)$, $X_t$ 为经过 PerMIF 算法变换的隐私保护图像;
输出: 隐私保护数据集 $P_t$、识别网络 $R_t$ 和转换网络 $T_t$ ($t=0,1,\ldots,K$).

1. **if** $t=0$ **do**
2. 在初始任务数据集 $D_0$ 上训练识别网络 $R_0$ 和转换网络 $T_0$
3. 通过 $T_0$ 将 $D_0$ 转换为隐私保护数据集 $P_0$, 并将 $P_0$ 保存到服务器
4. **else**
5. 在当前任务数据集 $D_t$ 上训练转换网络 $T_t$
6. 通过 $T_t$ 将 $D_t$ 转换为隐私保护数据集 $P_t$, 并将 $P_t$ 保存到服务器
   // $t'$ 表示到目前为止所有任务的编号 (包括当前任务)
7. 通过隐私保护数据集 $P_{t'}$ 训练当前任务识别网络 $R_t$
8. **end if**
9. **return** $P_t$、$R_t$ 和 $T_t$

本方法将前阶段任务的隐私保护模板进行保存, 在当前任务模型训练时重放以往任务的隐私保护模板, 与当前任务转换的隐私保护模板一起进行识别网络的扩展训练, 从而实现安全的可扩展虹膜识别, 识别网络的扩展学习过程如图 5 所示. 值得注意的是, PTR-SIR 方法中存储的所有先前任务的隐私保护模板均经过 TNCB 保护, 而并非原始训练数据. 识别网络具体的扩展训练过程如图 5 所示, 模型训练之前, 先将原始虹膜图像通过 PerMIF 算法进行处理, 再通过转换网络进行变换得到隐私保护模板作为当前任务的训练数据. 再将当前任务的训练数据与之

前任务所保存的训练数据结合起来同时训练识别网络, 通过公式 (2) 计算分类器输出的预测标签 $\hat{y}$ 和真实标签 $y$ 之间的分类损失 $L_{\text{class}}$ 以及公式 (9) 计算中心损失, 并根据分类损失和中心损失通过反向传播调整参数, 进而完成整个识别网络的扩展训练.

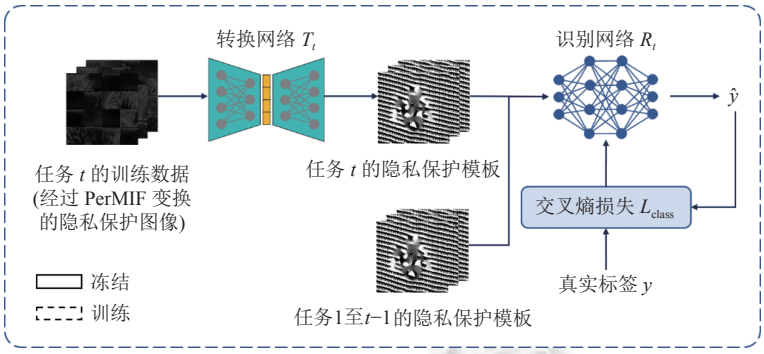


图 5 基于隐私保护模板重放的安全虹膜识别方法模型增量训练过程

## 4 安全性分析及效率分析

### 4.1 安全性分析

(1) 不可逆性. 本文所提出的两种可扩展安全虹膜识别方案都通过 TNCB 方法对生物数据进行保护. 首先对 TNCB 的不可逆性进行分析. TNCB 的不可逆变换主要包括两个阶段, 在通过 PerMIF 算法进行的第 1 阶段变换过程中, 不可逆变换包括取模和正反合并运算, 首先, 对于取模运算, 设经过分块置换之后的图像为 $P$, 取模运算的阈值为 $\tau$. 对于原始虹膜图像的任意值 $P_{i,j}$, 其中, $i \in [0,H]$, $j \in [0,W]$, $P_{i,j}$ 的值在 0–255 之间, $W$ 和 $H$ 分别为经过预处理的虹膜图像的宽和高. 设经过取模运算的图像为 $M$, 其中, $M_{i,j} = P_{i,j} \bmod \tau$, 由于取模运算在运算之后只保留了余数, 而商的信息被丢弃, 因此, 这一步运算是多对一映射. 经过取模运算之后, 对于每一个值 $M_{i,j}$, 都有 $[256/\tau]$ 种可能的值与之对应. 对于正反合并运算, 通过计算得到正反合并之后的图像 $X$, 其中, 正反合并的阈值为 $\gamma$, 通常将其值设为 $\tau/2-1$, 因此, $X_{i,j} \in [0,\gamma]$. 由于正反合并运算也是多对一映射, 对于正反合并之后的每一个值 $X_{i,j}$, 都对应存在两个可能的 $M_{i,j}$ 值. 因此, 对于一张经过取模和正反合并运算的图像 $X$, 对应可能的图像 $P$ 的数量为:

$$G = \left(2 \times \left\lfloor \frac{256}{\tau} \right\rfloor\right)^{W \times H} \tag{11}$$

从上述计算公式可以看出, 当 $\tau$ 越大时, 得到的多对一映射的数量越小, 当 $\tau$ 取最大值 128 时, 得到的 $G$ 值最小. 对于本文讨论的虹膜图像, 虹膜图像的长宽都为 128 个像素值, $G$ 的取值范围为:

$$G \geqslant \left(2 \times \left\lfloor \frac{256}{128} \right\rfloor\right)^{128 \times 128} = 2^{32768} \tag{12}$$

综合以上分析可知, 对于攻击者来说, 由经过 PerMIF 算法转换得到的图像恢复出取模运算之前的图像在计算上是非常困难的. 在第 2 阶段通过 U-Net 进行变换的过程也是不可逆变换, 在编码阶段主要的不可逆变换为卷积、池化和非线性激活 3 种运算. 卷积过程将输入矩阵与卷积核进行逐位相乘并求和得到卷积之后的结果. 假设输入图像的宽和高分别为 $W_{\text{in}}$ 和 $H_{\text{in}}$, 卷积核尺寸分别为 $W_k$ 和 $H_k$, 卷积步长为 $s$, 填充为 $p$, 经过卷积之后的输出图像宽和高分别为:

$$W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} - W_k + 2 \times p}{s} + 1 \right\rfloor \tag{13}$$

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} - H_k + 2 \times p}{s} + 1 \right\rfloor \tag{14}$$

根据上述计算方法可以看出, 在经过卷积运算之后矩阵的尺寸有所变小, 因此该运算存在多对一映射, 在卷积核不可逆的情况下, 该卷积运算是不可逆的. 此外, 在完成每一次卷积运算之后都有一次激活函数的运算, 该激活函数为非线性激活函数, 计算方法为:

$$ReLU(x)=\max(x,0)=\begin{cases}0, & x\leqslant 0\\ x, & x>0\end{cases} \tag{15}$$

结合上述公式可以发现, *ReLU* 激活函数在计算过程中也存在多对一映射, 对于所有的负输入, *ReLU* 函数都会输出 0, 这意味着多个不同的输入会映射到相同的输出, 存在信息丢失. 因此, 当激活函数的输出结果为 0 时, 无法区分输入是 0 还是任何一个负值, 所以该计算过程也是不可逆的.

此外, 在每一组卷积和激活函数计算完成之后, 该模型进行了一次最大池化运算, 最大池化在每一次计算时选择局部范围内的最大值进行输出, 对其他值进行丢弃, 目的是选择局部范围内最具代表性的特征. 不难看出, 在这一步运算过程中也丢失了很多信息, 因此这一步的计算也是不可逆的, 根据最大池化之后的结果无法完整恢复出池化之前的矩阵数据.

而在 U-Net 网络解码的阶段, 为了转化得到更大尺寸的图像, 对低维图像进行了上采样, 上采样采用最近邻插值实现, 在插值过程中往往也会引入误差. 并且解码过程中也伴随着卷积、池化和 *ReLU* 非线性激活函数的计算, 同样是不可逆运算. 通过公式 (10) 中的转换网络损失函数可知, 在转换网络训练过程中通过对抗训练最大化转换前后图像的特征损失, 因此训练过程中会不断扩大丢失的信息和引入的误差. 因此, 通过转换网络转换得到的图像无法重建图像出转换之前的图像.

根据上述讨论, TNCB 方法能够满足不可逆性. 接下来对 GFR-SR 和 PTR-SIR 方法的不可逆性进行讨论. 对于 GFR-SIR 方法, 该算法对图像同样采用 PerMIF 算法进行处理, 与上述 TNCB 对应部分不可逆性相同, 而对于第 2 阶段的不可逆变换, GFR-SIR 中对于 PerMIF 算法转换的图像先通过特征提取器进行特征提取, 再通过转换网络对该特征向量进行不可逆变换生成最终的隐私保护模板. 本方法中转换网络通过一维卷积、激活函数和池化等运算实现不可逆变换, 这些变换过程都存在多对一映射, 无法通过转换之后的结果完整恢复出转换之前的信息. 同样的, 由于采用了对抗的思想对转换网络进行训练, 在训练过程中会逐渐增大上述不可逆变换过程带来的信息损失, 因此, 本方法能够满足不可逆性. 而对于 PTR-SIR 方法, 该方法中是将 TNCB 方法转换的隐私保护模板进行了保存和重放, 因此该方法与 TNCB 方法同样满足不可逆性.

(2) 可撤销性. 可撤销性要求当某个用户的隐私保护模板发生泄露时, 该用户的模板能够被撤销, 无法继续用于虹膜识别或者认证, 与此同时, 用户还可以重新注册新的模板用于后续的虹膜识别. 对于 GFR-SIR, 当用户的模板发生泄露时, 首先将已经部署的识别网络和转换网络删除, 使其无法继续提供服务. 由于训练数据在完成模型训练之后已经被删除, 并未在服务器中进行保存. 因此需要重新收集用户的虹膜图像用于模型训练, 重新收集的图像通过修改 PerMIF 算法中的置换参数进行生成. 并完成后续模型训练, 最终将得到全新的转换网络和识别网络进行部署以提供虹膜识别服务. 对于 PTR-SIR, 当用户的模板发生泄露时, 由于该方案存储了多个转换网络, 仅需要将识别网络与该用户对应的转换网络进行删除, 同时也仅需收集该转换网络对应的用户虹膜图像用于转换网络的训练. 重新收集的图像通过修改 PerMIF 算法中的置换参数进行生成. 对于识别网络, 由于隐私保护模板进行了存储, 因此仅需结合泄露用户新的对应隐私保护模板数据进行识别网络的训练. 综合上述分析, 本文所提出的两种方案都满足可撤销性.

(3) 不可链接性. 不可链接性要求在不同应用中注册的虹膜模板之间不能进行交叉匹配, 即无法判断在不同应用中的模板是否来自同一用户. 对于本文的两种方法, 在 PerMIF 阶段执行了一次分块置换操作, 该分块置换通过随机生成的置换参数实现. 在不同的应用中, 即使相同的虹膜图像经过随机置换之后的图像也相当于是随机分布的. 所以在后续的取模、正反合并和转换网络变换之后, 得到的隐私保护图像也相当于是随机分布的. 因此, 来自不同应用之间的隐私保护虹膜模板之间进行交叉匹配时, 攻击者无法通过模板之间的相似度来判断两个模板是否来自相同的用户. 此外, 本文所得到的模板与 TNCB 方案相同, TNCB 方案的不可链接性已经得到证明. 综上所述, 本文提出的方法能够有效满足不可链接性.

### 4.2 效率分析

本文方法是在 TNCB 方法基础上进行的改进和扩展, 第 1 阶段通过 PerMIF 算法对原始虹膜图像进行转换的过程与 TNCB 中保持一致. 具体来说, 本方法对原始虹膜图像进行了分块置换、取模和正反合并运算等变换, 以生成经过初步隐私保护变换的虹膜图像用于模型训练. 文中涉及的虹膜图像都为单通道的灰度图像, 设每张经过预处理的虹膜图像的宽和高分别为 $W$ 和 $H$, 下面将展开分析本方法在模板转化过程中的效率.

首先, 对于分块置换操作, 需要先生成应用特定的置换参数. 方法中的分块置换的块大小为 $b$, 因此虹膜图像在宽方向上可以被划分为 $\lfloor W/b \rfloor$ 个块, 而在高方向上可以划分为 $\lfloor H/b \rfloor$ 个块. 由于本文所使用的虹膜图像宽和高相同, 因此将两个方向上划分的块数都记为 $m$. 而生成随机置换参数的方法为对位置坐标矩阵分别按列和行执行洗牌算法, 因此生成置换参数的时间复杂度为 $O(m^2)$. 接下来根据生成的置换参数对原始虹膜图像逐块执行置换操作, 时间复杂度也为 $O(m^2)$. 因此整个分块置换操作的时间复杂度为 $O(m^2)$. 接下来, 本方法对于置换得到的虹膜图像执行逐像素的取模和正反合并运算, 这两个运算的时间复杂度都为 $O(W \times H)$, 因此这个过程整体的时间复杂度为 $O(W \times H)$.

假设数据集中虹膜图像的数量为 $n$, 在模型训练前需对数据集中的每一张图像都按照上述算法进行变换, 因此整个数据集变换的时间复杂度为 $O(n \times m^2) + O(n \times W \times H)$. 由于参数 $W$ 和 $H$ 为常数, 因此模板转换过程整体的时间复杂度为 $O(W \times H)$. 下面将重点分析本文提出的两种安全增量虹膜识别方法在模型增量训练过程中的效率.

(1) 数据存储和处理效率. GFR-SIR 在增量学习阶段通过当前任务的数据和生成器重放的特征进行模型训练, 生成特征的过程需要一些额外的计算资源, 但避免了直接存储和处理大量历史数据, 在数据存储和存储效率上具有一定优势, 但该方法需要训练和存储生成器模型, 对计算资源和存储效率也有一定的影响. 而 PTR-SIR 在每一次增量学习过程中都需要同时处理以往任务的所有隐私保护模板和当前任务转换的模板, 处理的数据量随时间呈线性增长, 会导致数据预处理和加载的时间增加, 从而影响整体的效率. 并且在大规模增量学习场景下将会给数据存储和计算资源带来很大的挑战, 因此该方法在数据存储和处理效率方面有一定的局限性.

(2) 模型更新计算效率. GFR-SIR 通过生成器重放以往任务的特征参与当前模型的训练, 相比于直接通过完整数据联合训练减少了每次迭代训练需要处理的数据量, 因此模型更新的计算效率较高. 但是, 该方法在每个任务中还需要另外训练一个特征生成器, 在生成器的对抗训练过程中也需要消耗一定的计算资源, 会对效率有一定的影响. 而 PTR-SIR 在增量学习过程中, 一方面是该方法首先需要先在当前任务数据集上进行一次识别网络和转换网络的训练, 用于将当前任务的数据转换为隐私保护模板, 这一过程需要消耗一定的计算资源. 但由于该模型的训练只是在当前任务数据集上进行计算, 因此不会显著影响方法的效率. 另一方面, 本方法需要在到目前为止所有任务的隐私保护模板上进行计算以更新模型参数, 这一过程显著增加了每次训练所需的计算资源和时间, 尤其是任务数量较多时, 需要消耗的计算资源更多, 效率也相对更低.

## 5 实验结果与分析

### 5.1 实验设置

实验在中国科学院的两个经典的虹膜数据集 CASIA-Iris-Lamp 和 CASIA-Iris-Thousand[34]上进行, 后文分别记为 Lamp 和 Thousand. 其中, Lamp 数据集中包含来自 411 个用户的 16 215 张虹膜图像, 该数据集是在不同光线条件下采集到的. 在不同光线下采集到的虹膜图像同一类内具有一定的变化, 从而可以更有效评估虹膜识别方法的有效性. Thousand 数据集中包括 2 000 张来自 1 000 个不同用户的虹膜图像, 该数据集中包括正常眼部图片以及戴眼镜有反光的图像, 由于镜面反射和灯光的影响, 该数据集的识别难度非常大, 常被用于评估虹膜识别方法的鲁棒性. 本文采用 Othman 等人[35]提出的 OSIRIS-V4.1 系统进行虹膜图像的定位、分割和归一化等操作之后得到归一化虹膜图像, 该图像的尺寸为 $64 \times 256$. 随后对其进行分割和拼接得到 $128 \times 128$ 的虹膜图像, 本文的所有实验都是基于该尺寸的图像展开的.

实验的初始任务类别数量和扩展任务数量的设置参照文献 [33] 中的方法, 具体来说, 对于 Lamp 数据集, 为了

便于进行增量任务划分, 本实验中选择左眼数据集的前 400 个类别进行模型的扩展任务实验, 将初始任务类别设置为 200, 由于在初始任务类别为 200 时剩下类别无法平均分为 3 个任务, 因此为了保持统一, 本文在实验中将剩下的类别分别按照任务数量为 1、2、4、5 平均划分为不同的任务进行扩展任务. 对于 Thousand 数据集, 则是通过完整的 1 000 个类别进行扩展任务, 初始任务类别数量设置为 500, 同样将剩下的类别根据不同的任务数量进行平均划分. 对这两个数据集, 本文以 4:1 的比例划分训练集与测试集.

在方法的具体实现中, 本文采用 U-Net[36]作为转换网络, 此外, 计算特征重建损失通过卷积神经网络 VGG16[37]进行计算, 该损失通过将转换前后的图像输入到 VGG16 之后, 计算第 2 个 *ReLU* 激活函数生成的特征图之间差异得到. 识别网络的主体架构为 Gangwar 等人[38]提出的专门用于虹膜识别的卷积神经网络 DeepIrisNet-A, 同时引入中心损失对 DeepIrisNet-A 进行改进, 识别网络的结构如图 6 所示. 使用中心损失改进的 DeepIrisNet-A 能够很好地提取出虹膜数据的特征, 对于虹膜数据有较高的适配性. GFR-SIR 与 PTR-SIR 方法中的参数设置与 TNCB 方法[31]中相同, 如表 1 所示.

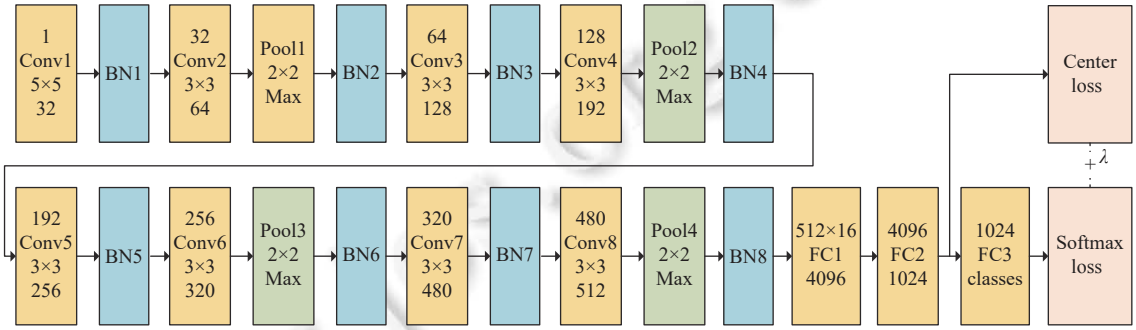


图 6 识别网络架构

表 1 实验参数设置

| 实验参数 | 参数设置 |
|---|---|
| 学习率 | 识别网络: 0.000 1; 中心损失: 0.001; 转换网络: 0.01 |
| 批量大小 | 32 |
| 块大小 | 32 |
| 训练轮次 | 200 |
| 中心损失和转换网络权重 | 1 |

对于安全可扩展虹膜识别方法的性能评估, 本文采用的指标是平均识别准确率以及遗忘率. 其中, 识别准确率通过计算到目前为止所有任务的平均识别准确率得到, 假设当前任务编号为 $k$, 本文将模型在第 $j$ 个任务的数据集上得到的识别准确率记为 $acc_{k,j}$. 与文献 [33] 中的计算方法一样, 本文在计算平均识别准确率时采用的是加权平均. 具体来说, 首先通过式 (16) 计算当前任务模型在每一任务数据集上的识别准确率, 再计算加权平均值得到最终的平均识别准确率, 计算方法为:

$$acc_k = \frac{\sum_{j=0}^{k}(w_j \times acc_{k,j})}{\sum_{j=0}^{k} w_j} \tag{16}$$

其中, $w_j$ 表示第 $j$ 个任务所占的权重, 在具体计算中可将其设置为第 $j$ 个任务的类别数量. 此外, 本文中还采用了 Chaudhry 等人[39]提出的平均遗忘 $f_j^k$ 这一指标来评估扩展任务过程中模型对于之前阶段任务知识的遗忘, 该指标将遗忘定义为任务 $j$ 在整个学习过程中得到的最高识别准确率与当前任务 $k$ 中实现的识别准确率的差值, 其值的范围为: $f_j^k \in [-1, 1]$, 计算方法为:

$$f_j^k = \max_{l\in\{j,\ldots,k-l\}} acc_{l,j} - acc_{k,j} \tag{17}$$

在计算得到任务 $k$ 之前每个任务的遗忘指标 $f_j^k$ 之后, 本文中同样通过加权平均的方式计算得到任务 $k$ 的平均遗忘 $F_k$, 计算方法如下:

$$F_k = \frac{\sum_{j=0}^{k-l}(w_j \times f_j^k)}{\sum_{j=0}^{k} w_j} \tag{18}$$

### 5.2 性能和遗忘评估

(1) Lamp 数据集上的识别准确率和遗忘率评估. 本节给出了在 Lamp 数据集上的识别准确率和遗忘率评估的实验结果, 与第 5.1 节不同的是, 在本节实验中将初始任务类别数量设置为 200, 实验结果如图 7 和图 8 所示. 首先, 从图 7 和图 8 的结果对比可以发现, 将初始任务设置为 200 时, 3 种方法都取得了更高的识别准确率. 这是由于初始任务类别数量更多则用于扩展训练的类别数量就更少, 在扩展任务过程中在新数据上进行训练对旧模型的参数修改也更少, 由此带来的知识遗忘也更少.

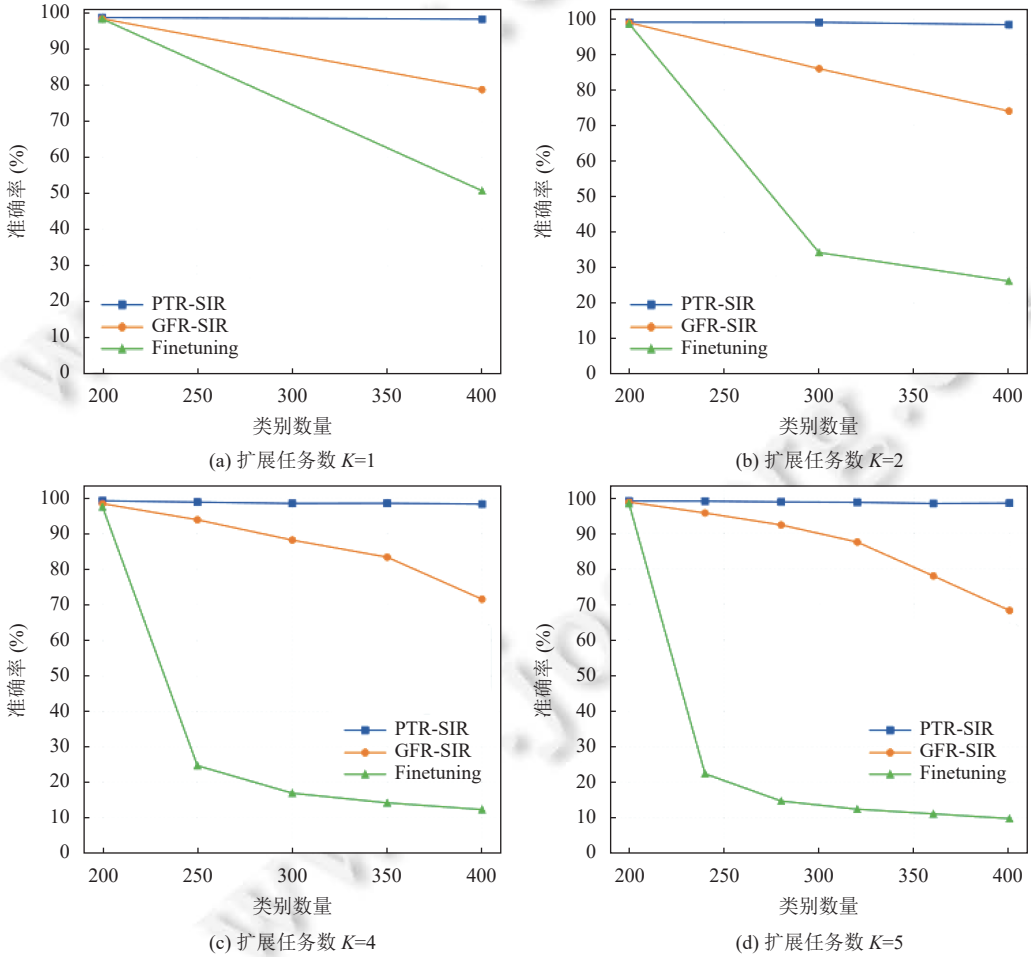


图 7 Lamp 数据集上的识别准确率评估实验结果

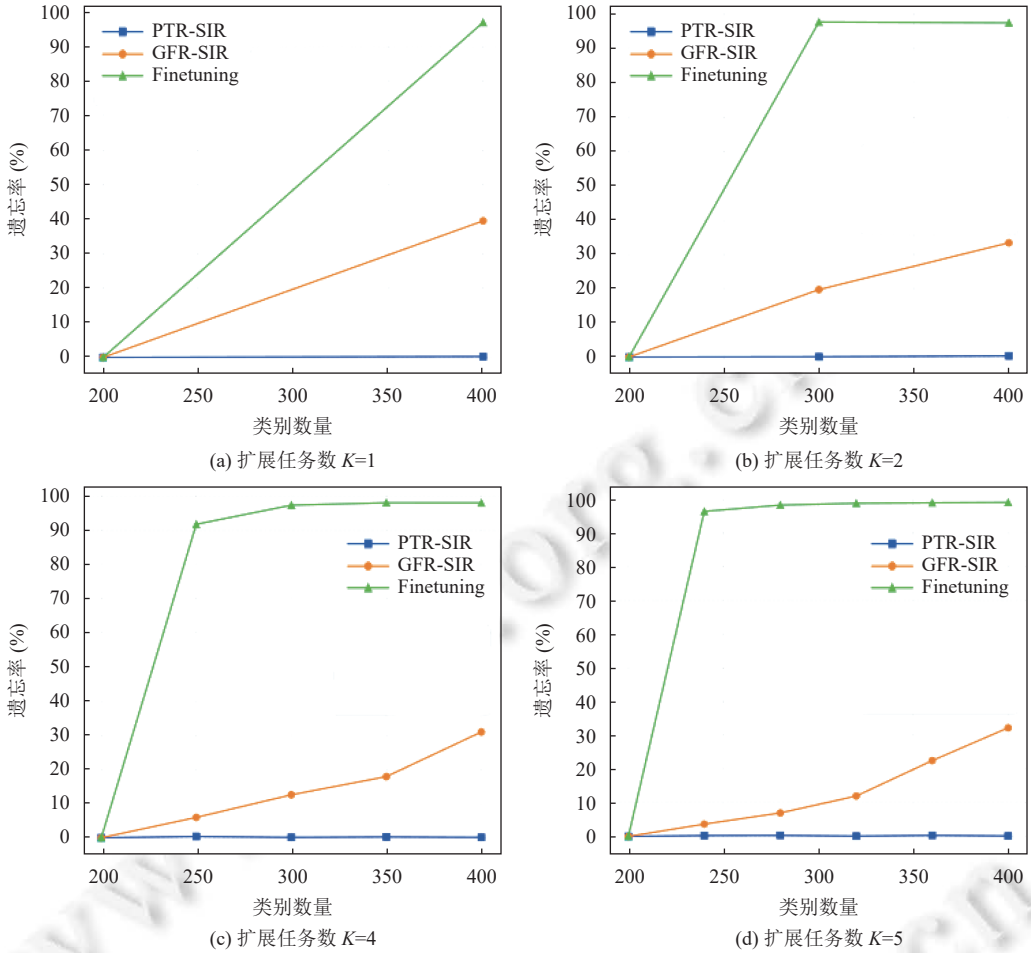


图 8 Lamp 数据集上的遗忘率评估实验结果

从整体趋势可以看出, 随着任务数量的增加, 3 个方法在完成最后一轮任务增量学习之后的识别准确率都有所下降, 但各个方法下降的幅度和下降趋势不同. 具体来说, 随着扩展任务数量的增加, Finetuning 方法的识别准确率下降趋势是先急剧下降到较低水平, 后续变化趋于平缓. 该方法的识别准确率下降较大的原因是该方法在新数据上微调模型, 从而导致模型主要学习到了当前任务的知识, 而遗忘了以往任务的大部分知识, 因此在以往任务数据集上的识别准确率显著下降. GFR-SIR 方法的变化趋势是前几个任务下降较为缓慢, 而后面几个任务下降速度变快, 尽管该方法在很大程度上缓解了遗忘, 降低了性能的下降幅度, 但在扩展任务过程中由于生成器生成的特征质量的影响, 还是有一定的知识遗忘, 并且随着扩展任务数量的增加, 知识的遗忘逐渐累积, 导致识别准确率下降逐渐加快. 而 PTR-SIR 方法的识别准确率变化趋势较为平缓, 相对于初始任务只有轻微下降, 这是因为 PTR-SIR 实现了在保护虹膜隐私数据的前提下让以往任务的数据以另一种形式参与到了当前任务模型的训练中.

遗忘率评估实验结果如图 8 所示, 从整体趋势上看, 不同方法的遗忘率曲线都是随着任务数量的增加而有所上升, 但 3 个方法变化的趋势有所不同. 其中, Finetuning 方法在完成 1 轮任务的扩展训练之后遗忘率迅速上升, 达到了 90% 以上, 而后趋于平缓, 也说明了该方法在扩展任务过程中的灾难性遗忘现象. GFR-SIR 方法的遗忘率曲线随着任务数量的增加先缓慢上升, 而后有所加快, 在完成 5 轮任务的扩展训练之后为 32.24%, 相对于微调方案在很大程度上缓解了知识的遗忘. 而 PTR-SIR 方法的遗忘率曲线则是趋于平缓, 在完成 5 轮任务的扩展训练之后遗忘率为 0.14%, 说明该方法在很大程度上缓解了扩展任务过程中的知识遗忘, 也从另一个角度说明了该方法在

缓解性能下降方面的有效性.

(2) Thousand 数据集上的识别准确率和遗忘率评估. 本节还给出了在 Thousand 数据集上的识别准确率和遗忘率评估实验结果, 实验结果如图 9 和图 10 所示. 从整体的趋势来看, 微调方案同样在完成一轮任务扩展训练后识别准确率就出现了大幅度下降, 而后下降幅度减小. GFR-SIR 方法则是在前几轮任务扩展训练时准确率下降幅度较小, 而后面几轮任务的识别准确率下降较大. 但从结果也可以看出, GFR-SIR 相对于微调方法在识别准确率方面有了较大提升. 此外, PTR-SIR 方法的识别准确率的变化幅度较小, 但任务数量大于 1 时的结果中都出现了先下降后上升的变化趋势.

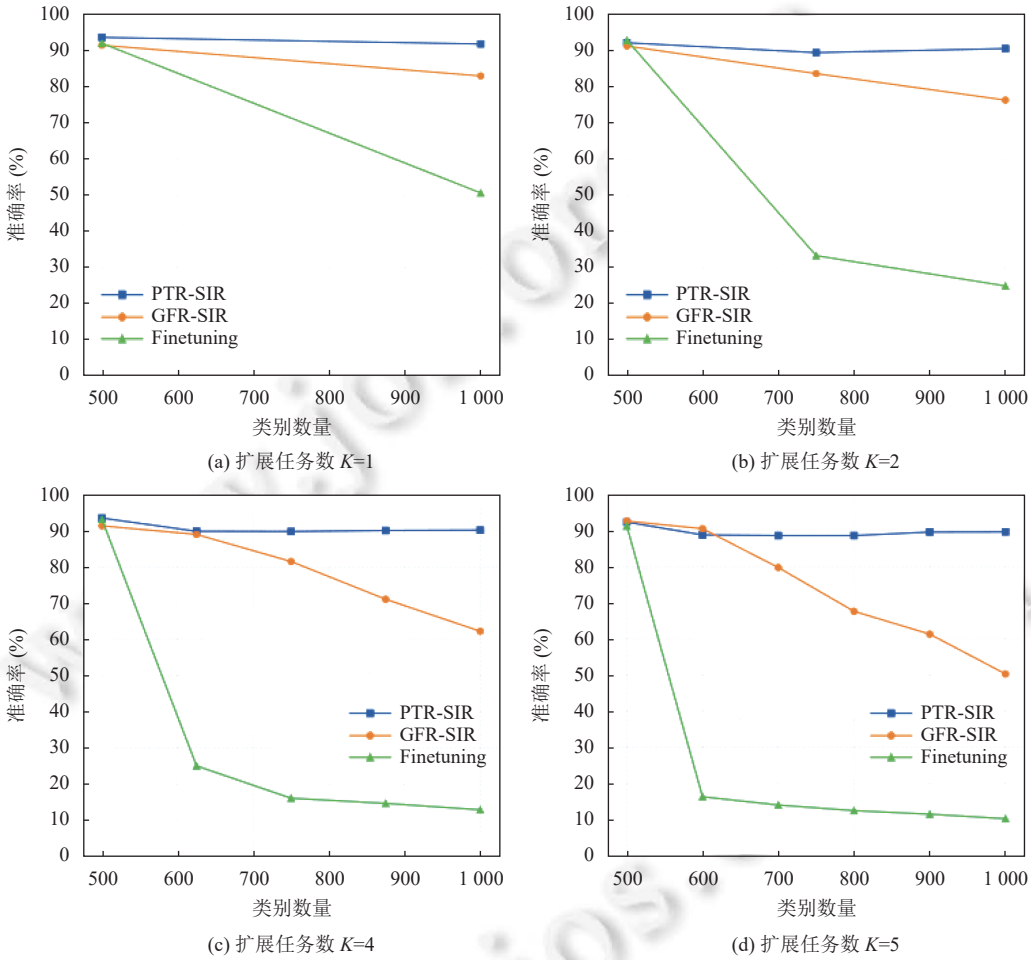


图 9 Thousand 数据集上的识别准确率评估实验结果

从具体的实验结果来看, 在该数据集上 3 个方法的第 1 轮任务上的识别准确率都在 92% 左右. 当扩展任务数为 5 时, 微调方法的性能下降幅度最大, 完成 5 轮任务模型训练时的识别准确率只有 10.21%, 下降了 81.02%. 而 GFR-SIR 方法的识别准确率为 50.27%, 和第 1 轮任务识别准确率相比下降了 40.33%. 尽管本方法相比于微调方案有了很大提升, 但与初始任务识别准确率相比仍有一定幅度的下降. 一方面是因为扩展任务的难度会随着任务规模的增大而增大, 而在 Thousand 数据集上总共有 1 000 个类别, 扩展任务的难度比较大. 其次, 由于该数据集中存在很多戴眼镜的人眼图片, 由于镜片的反射和光线的影响, 使得该数据集的学习难度很大, 因此在增量学习中精度下降也会较大. 最后, 该数据集中每个类别的虹膜图片只有 10 张, 在模型训练过程中也存在模型训练不充分的

问题. 而 PTR-SIR 方法在完成 5 轮的扩展任务之后性能为 89.6%, 相对于第 1 轮的性能下降了 2.72%, 也说明了该方法在保持性能方面的有效性. 此外, 从实验结果还发现一个现象, 在任务数量为 5 时, 在完成第 1 轮的扩展任务时出现了 GFR-SIR 方法的识别准确率比 PTR-SIR 方法更高. 经过分析可知, 这是由于在该参数设置下, 第 1 轮扩展任务过程中通过 PerMIF 变换和转换网络将虹膜图像转换为隐私保护模板造成的特征损失较大, 从而导致在生成的隐私保护模板上训练的识别网络的识别准确率较低, 本文认为造成该任务下隐私保护模板特征损失较大与该任务所对应类别数据的特性有关.

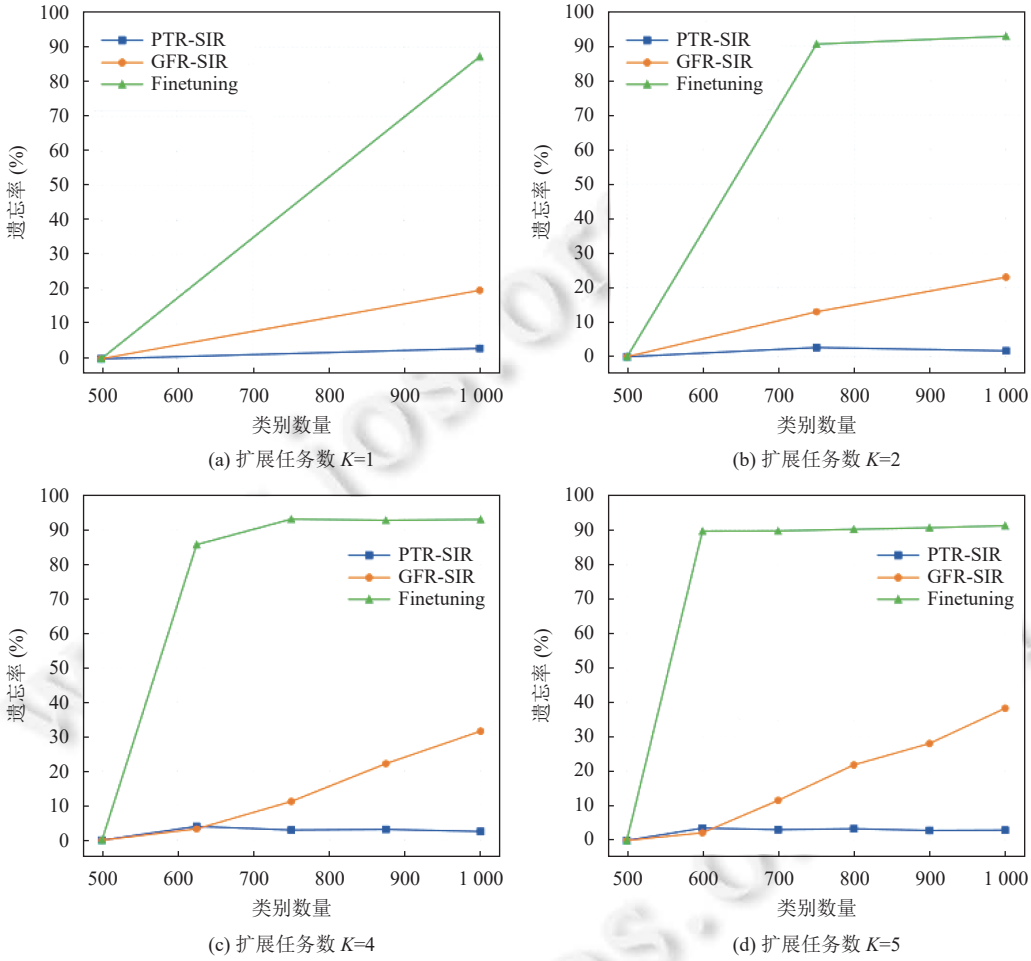


图 10 Thousand 数据集上的遗忘率评估实验结果

此外, 从图 10 中的遗忘率评估实验结果可以看出, 微调方案的遗忘率曲线变化规律与 Lamp 上类似, 在完成 1 轮任务扩展训练之后模型的遗忘率就达到了 80% 以上, 而完成 5 轮任务的扩展训练时遗忘率为 91.19%. 本文的 GFR-SIR 方案遗忘率的变化趋势则是随着任务数量的增加而增加, 在完成 5 轮任务的扩展训练之后遗忘率接近 40%, 说明本方法相对于微调的方法在很大程度上缓解了知识的遗忘. 而 PTR-SIR 方法的遗忘率变化趋势则是先上升后下降, 该变化趋势的原因已在识别准确率评估部分进行了分析. 在完成 5 轮任务的扩展训练时, 遗忘率接近 3%, 相对于微调方法降低超过 88%, 该结果也说明了本方法可以更加有效地缓解知识遗忘. 从图 9 中的识别准确率实验结果和图 10 中的遗忘率实验结果对比可以发现, 识别准确率和遗忘率的变化呈现相反的趋势, 说明本文方法在减少扩展任务识别准确率降低幅度的同时, 也降低了遗忘率提升的幅度. 因此, 两个方面的评估结果相互验

证, 都说明了本文提出的方法在缓解遗忘和保持性能方面的有效性.

### 5.3 消融实验

本节探究了 GFR-SIR 方法的不同模块对于识别准确率的影响, GFR-SIR 主要由特征蒸馏模块和生成特征重放模块组成, 本文设置了 4 种方法进行消融实验: 微调 (不使用任何模块)、特征蒸馏、生成特征重放、GFR-SIR (同时使用特征蒸馏模块和生成特征重放模块).

消融实验结果如表 2 所示, 可以看出在微调方法下识别准确率下降较为显著, 在完成第 1 轮任务的扩展训练之后, 识别准确率下降到了 22.42%, 说明该方法对于之前学习到的知识的遗忘非常严重. 特征蒸馏方法下识别准确率下降得到了缓解, 在完成第 1 轮任务的扩展训练之后, 识别准确率仍然达到 86.94%, 相对于模型微调的方法, 该方法的识别准确率在每轮都有了一定的提升. 此外, 从实验结果还可以看出, 随着任务数量的增加, 该方法的识别准确率下降也越来越大, 在完成 5 轮扩展任务时的识别准确率只有 15.36%. 生成特征重放方法下, 相对于微调方法和特征蒸馏的方法性能都有了一定的提升, 在完成 5 轮的扩展任务时识别准确率为 34.01%, 说明相对于特征蒸馏模块, 生成特征重放模块在缓解遗忘方面更有效. 最后, GFR-SIR 方法相对于前面的 3 种方法, 在完成各轮的扩展任务之后都实现了更高的识别准确率, 说明 GFR-SIR 方法在缓解遗忘方面的有效性. 相对于只有其中一个模块的方法, 本方法识别准确率提升更大, 本文认为这是因为特征蒸馏模块只缓解了特征提取器的遗忘, 而生成特征重放则是缓解了分类器的遗忘.

表 2 消融实验结果 (%)

| 方法 | 增量任务编号 | | | | | |
|---|---|---|---|---|---|---|
| | T0 | T1 | T2 | T3 | T4 | T5 |
| 微调 | 98.39 | 22.42 | 14.75 | 12.46 | 11.15 | 9.83 |
| 特征蒸馏 | 98.91 | 86.94 | 69.44 | 49.08 | 28.49 | 15.36 |
| 生成特征重放 | 98.73 | 88.98 | 78.12 | 64.64 | 45.29 | 34.01 |
| GFR-SIR | 98.80 | 95.73 | 92.37 | 87.55 | 78.03 | 68.32 |

### 5.4 不同方法的性能对比

本节将给出所提出的方案与其他不同增量学习方法的识别准确率对比结果, 所有方法的对比结果都是在 Lamp 数据集上将初始任务类别数量设置为 200, 剩下的类别数量按照划分为 5 轮任务设置. 接下来将对对比的方法的设置进行介绍. 明文数据联合训练方法在每轮任务模型训练过程中都通过之前所有任务的明文数据进行联合训练. 模型微调方法与消融实验中的微调方法设置相同, 剩余的经典增量学习方法 (LwF[40]、iCaRL[41]、EWC[42]、PODNet[43]、WA[44]、DER[45]、FOSTER[46]、FeTrIL[47]) 使用开源框架 PyCIL[48]实现, 均采用 ResNet32 网络作为骨干框架, 并分别使用明文数据和隐私保护数据作为训练数据进行扩展任务训练. 尽管在前边的讨论中提到在扩展任务训练过程中, 考虑到生物数据隐私泄露的问题, 不应该保存明文生物数据, 但是部分经典增量学习方法必须保存部分先前任务的数据才能进行, 因此本文对这些增量学习方法中每个类都保留一个样本以达到这些方法的最低需求.

性能对比实验结果如表 3 所示, 明文数据联合训练的方法的性能几乎没有下降, 在完成 5 轮任务的扩展训练之后识别准确率仍然有 99.48%. 基于模型微调的方法在完成 5 轮任务的扩展训练之后的识别准确率为 9.83%. 此外, 表 3 中还给出了与另外 8 种经典的增量学习方法 LwF、iCaRL、EWC、PODNet、WA、DER、FOSTER、FeTrIL 在分别采用明文数据和受保护数据训练的情况下的对比结果. 其中, LwF、EWC 方法在明文数据情况下完成 5 轮任务的扩展训练后识别准确率均下降幅度较大, 而 iCaRL、PODNet、WA、DER、FOSTER、FeTrIL 方法在明文数据情况下完成 5 轮任务的扩展训练后识别准确率下降幅度较小. 然而在采用受保护数据进行训练的情况下, 对比的 8 种方法识别准确率均大幅度下降, 最高为 DER 方法达到的 54.86%. 本文提出的 GFR-SIR 方法在各轮任务的识别准确率优于采用受保护数据进行训练的现有方法. 尽管在某些情况下, GFR-SIR 方法的准确率低于部分采用明文数据训练的方案, 但 GFR-SIR 方法在无须存储任何原始生物数据的情况下, 仍能够有效完成扩展训

练任务. 这不仅实现了较高的识别准确率, 还显著提高了模型的安全性. 此外, 从实验结果可以看出, 本文提出 PTR-SIR 方法的性能保持得更好, 在完成 5 轮任务的增量学习之后的识别准确率仍保持在 98.49%, 高于对比的其他方案, 但 PTR-SIR 方法由于需要为每次任务单独训练转换网络并存储, 相比 GFR-SIR 方法更加复杂, 成本更高, 因此更适合对精度需求高的场景. 综上所述, 本文提出的两种方案在保证生物数据隐私安全的情况下实现了较高的准确率完成扩展训练任务.

表 3　不同方法识别准确率对比 (%)

| 方法 | 训练数据 | 增量任务编号 | | | | | |
|---|---|---|---|---|---|---|---|
| | | T0 | T1 | T2 | T3 | T4 | T5 |
| 联合训练 | 明文 | 99.70 | 99.68 | 99.64 | 99.54 | 99.56 | 99.48 |
| 模型微调 | 明文 | 98.39 | 22.42 | 14.75 | 12.46 | 11.15 | 9.83 |
| LwF[40] | 明文 | 97.85 | 85.71 | 64.93 | 48.10 | 39.93 | 33.58 |
| | 受保护数据 | 70.93 | 44.70 | 27.40 | 19.96 | 15.86 | 14.48 |
| EWC[42] | 明文 | 97.98 | 51.58 | 31.94 | 21.15 | 16.87 | 15.08 |
| | 受保护数据 | 71.18 | 40.01 | 20.00 | 13.96 | 10.07 | 8.05 |
| iCaRL[41] | 明文 | 97.85 | 95.07 | 90.15 | 86.49 | 84.54 | 81.20 |
| | 受保护数据 | 70.93 | 55.17 | 38.90 | 35.15 | 30.82 | 26.43 |
| PODNet[43] | 明文 | 98.61 | 96.02 | 90.97 | 87.58 | 84.64 | 82.64 |
| | 受保护数据 | 65.66 | 53.23 | 36.71 | 29.26 | 25.28 | 20.17 |
| WA[44] | 明文 | 97.85 | 93.36 | 89.01 | 86.59 | 85.07 | 84.09 |
| | 受保护数据 | 70.93 | 66.13 | 54.99 | 51.43 | 48.93 | 45.90 |
| DER[45] | 明文 | 98.86 | 96.03 | 94.09 | 94.87 | 93.90 | 93.63 |
| | 受保护数据 | 71.18 | 73.21 | 63.79 | 61.01 | 59.27 | 54.86 |
| FOSTER[46] | 明文 | 98.86 | 94.06 | 88.56 | 84.07 | 82.86 | 71.89 |
| | 受保护数据 | 70.18 | 48.69 | 33.16 | 29.94 | 29.67 | 25.99 |
| FeTrIL[47] | 明文 | 98.23 | 97.22 | 95.81 | 94.84 | 94.84 | 94.83 |
| | 受保护数据 | 68.80 | 54.58 | 50.22 | 46.48 | 42.31 | 40.83 |
| GFR-SIR | 受保护数据 | 98.80 | 95.73 | 92.37 | 85.55 | 78.03 | 68.32 |
| PTR-SIR | 受保护数据 | 99.12 | 99.06 | 98.83 | 98.74 | 98.39 | 98.49 |

## 6 总　结

针对虹膜识别模型在扩展过程中的灾难性遗忘, 进而导致识别准确率显著下降的问题, 本文对基于转换网络的安全虹膜识别方法进行了改进和扩展, 提出了两种基于隐私保护数据重放安全可扩展虹膜识别方法, 分别为基于生成特征重放的安全可扩展虹膜识别方法 GFR-SIR 和基于隐私保护模板重放的安全可扩展虹膜识别方法 PTR-SIR. 其中, GFR-SIR 方法引入了生成特征重放和特征蒸馏两种方法缓解增量学习过程中的知识遗忘. 与 TNCB 方法不同的是, 由于在当前阶段是通过重放特征参与分类器的训练, 因此本方法中通过特征转换网络对特征进行不可逆变换以生成隐私保护模板, 转换网络同样是通过对抗训练得到的. PTR-SIR 方法则是将以往任务的训练数据转换为隐私保护模板之后进行保存, 在训练当前任务的识别网络时重放隐私保护模板参与当前任务识别网络的训练, 从而缓解知识的遗忘. 在虹膜数据集上的实验结果表明, 所提出的两种方案在保护数据隐私的同时, 有效缓解了知识遗忘现象, 并显著降低了扩展任务中识别准确率的损失. 由于 GFR-SIR 方法不保存先前任务的训练数据, 该方法展现出更高的安全性和效率, 同时降低了内存需求, 尤其适用于内存资源有限的应用场景. 相比之下, PTR-SIR 方法在保持增量学习识别准确率方面表现更佳, 其准确率相较于明文数据联合训练的方法仅下降了 0.99%, 因此更适用于对识别准确率要求较高且具备充足内存的应用场景. 在未来的工作中, 考虑将提出的两种方案结合成一种适用于更多生物识别方法的安全可扩展框架.

**References:**


[1] Song JW, Cheng DF, Wang WF. Biometrics and Standardization. Beijing: China Electronics Standardization Institute, 2023. 42–53 (in Chinese).

[2] Al-Raisi AN, Al-Khouri AM. Iris recognition and the challenge of homeland and border control security in UAE. Telematics and Informatics, 2008, 25(2): 117–132. [doi: 10.1016/j.tele.2006.06.005]

[3] ISO/IEC. ISO/IEC 24745: 2011 Information technology—Security techniques—Biometric information protection. Geneva: Int'l Organization for Standardization, 2011.

[4] Hermans J, Mennink B, Peeters R. When a bloom filter is a doom filter: Security assessment of a novel iris biometric template protection system. In: Proc. of the 2014 Int'l Conf. of the Biometrics Special Interest Group. Darmstadt: IEEE, 2014. 1–6.

[5] Ghammam L, Karabina K, Lacharme P, Thiry-Atighehchi K. A cryptanalysis of two cancelable biometric schemes based on index-of-max hashing. IEEE Trans. on Information Forensics and Security, 2020, 15: 2869–2880. [doi: 10.1109/TIFS.2020.2977533]

[6] Dong XB, Jin Z, Jin ATB. A genetic algorithm enabled similarity-based attack on cancellable biometrics. In: Proc. of the 10th IEEE Int'l Conf. on Biometrics Theory, Applications and Systems. Tampa: IEEE, 2019. 1–8. [doi: 10.1109/BTAS46853.2019.9185997]

[7] Ouda O, Chaoui S, Tsumura N. Security evaluation of negative iris recognition. IEICE Trans. on Information and Systems, 2020, E103.D(5): 1144–1152. [doi: 10.1587/transinf.2019EDP7276]

[8] Ignatenko T, Willems FMJ. Information leakage in fuzzy commitment schemes. IEEE Trans. on Information Forensics and Security, 2010, 5(2): 337–348. [doi: 10.1109/TIFS.2010.2046984]

[9] Simoens K, Tuyls P, Preneel B. Privacy weaknesses in biometric sketches. In: Proc. of the 30th IEEE Symp. on Security and Privacy. Oakland: IEEE, 2009. 188–203. [doi: 10.1109/SP.2009.24]

[10] Rathgeb C, Uhl A. Statistical attack against iris-biometric fuzzy commitment schemes. In: Proc. of the 2011 CVPR WORKSHOPS. Colorado Springs: IEEE, 2021. 23–30. [doi: 10.1109/CVPRW.2011.5981720]

[11] Blanton M, Aliasgari M. Analysis of reusability of secure sketches and fuzzy extractors. IEEE Trans. on Information Forensics and Security, 2013, 8(9): 1433–1445. [doi: 10.1109/TIFS.2013.2272786]

[12] Cavoukian A, Stoianov A. Biometric encryption. Biometric Technology Today, 2007, 15(3): 11. [doi: 10.1016/S0969-4765(07)70084-X]

[13] Rathgeb C, Uhl A. A survey on biometric cryptosystems and cancelable biometrics. EURASIP Journal on Information Security, 2011, 2011(1): 3. [doi: 10.1186/1687-417X-2011-3]

[14] Juels A, Sudan M. A fuzzy vault scheme. Designs, Codes and Cryptography, 2006, 38(2): 237–257. [doi: 10.1007/s10623-005-6343-z]

[15] Juels A, Wattenberg M. A fuzzy commitment scheme. In: Proc. of the 6th ACM Conf. on Computer and Communications Security. Singapore: ACM, 1999. 28–36. [doi: 10.1145/319709.319714]

[16] Dodis Y, Reyzin L, Smith A. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In: Proc. of the 2004 Int'l Conf. on the Theory and Applications of Cryptographic Techniques. Interlaken: Springer, 2004. 523–540. [doi: 10.1007/978-3-540-24676-3_31]

[17] Chang EC, Shen R, Teo FW. Finding the original point set hidden among chaff. In: Proc. of the 2006 ACM Symp. on Information, Computer and Communications Security. Taipei: ACM, 2006. 182–188. [doi: 10.1145/1128817.1128845]

[18] Poon HT, Miri A. A collusion attack on the fuzzy vault scheme. The ISC Int'l Journal of Information Security, 2009, 1(1): 27–34. [doi: 10.22042/isecure.2015.1.1.4]

[19] Scheirer WJ, Boult TE. Cracking fuzzy vaults and biometric encryption. In: Proc. of the 2007 Biometrics Symp. Baltimore: IEEE, 2007. 1–6. [doi: 10.1109/BCC.2007.4430534]

[20] Kelkboom EJC, Breebaart J, Kevenaar TAM, Buhan I, Veldhuis RNJ. Preventing the decodability attack based cross-matching in a fuzzy commitment scheme. IEEE Trans. on Information Forensics and Security, 2011, 6(1): 107–121. [doi: 10.1109/TIFS.2010.2091637]

[21] Rathgeb C, Breitinger F, Busch C. Alignment-free cancelable iris biometric templates based on adaptive bloom filters. In: Proc. of the 2013 Int'l Conf. on Biometrics. Madrid: IEEE, 2013. 1–8. [doi: 10.1109/ICB.2013.6612976]

[22] Jin Z, Hwang JY, Lai YL, Kim S, Teoh ABJ. Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing. IEEE Trans. on Information Forensics and Security, 2018, 13(2): 393–407. [doi: 10.1109/TIFS.2017.2753172]

[23] Sadhya D, Raman B. Generation of cancelable iris templates via randomized bit sampling. IEEE Trans. on Information Forensics and Security, 2019, 14(11): 2972–2986. [doi: 10.1109/TIFS.2019.2907014]

[24] Lai YL, Jin Z, Teoh ABJ, Goi BM, Yap WS, Chai TY, Rathgeb C. Cancellable iris template generation based on indexing-first-one hashing. Pattern Recognition, 2017, 64: 105–117. [doi: 10.1016/j.patcog.2016.10.035]

[25] Zhao DD, Luo WJ, Liu R, Yue LH. Negative iris recognition. IEEE Trans. on Dependable and Secure Computing, 2018, 15(1): 112–125. [doi: 10.1109/TDSC.2015.2507133]

[26] Abdellatef E, Soliman RF, Omran EM, Ismail NA, Elrahman SESA, Ismail KN, Rihan M, Amin M, Eisa AA, El-Samie FEA. Cancelable face and iris recognition system based on deep learning. Optical and Quantum Electronics, 2022, 54(11): 702. [doi: 10.1007/s11082-022-03770-0]

[27] Kim J, Jung YG, Teoh ABJ. Multimodal biometric template protection based on a cancelable SoftmaxOut fusion network. Applied Sciences, 2022, 12(4): 2023. [doi: 10.3390/app12042023]

[28] Pandey RK, Zhou YB, Kota BU, Govindaraju V. Deep secure encoding for face template protection. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition Workshops. Las Vegas: IEEE, 2016. 77–83. [doi: 10.1109/CVPRW.2016.17]

[29] Chen LY, Zhao GH, Zhou JW, Ho ATS, Cheng LM. Face template protection using deep LDPC codes learning. IET Biometrics, 2019, 8(3): 190–197. [doi: 10.1049/iet-bmt.2018.5156]

[30] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: Proc. of the 14th European Conf. Amsterdam: Springer, 2016. 694–711. [doi: 10.1007/978-3-319-46475-6_43]

[31] Zhao DD, Liao HC, Liao SS, Li HH, Xiang JW. Cancelable iris biometrics based on transformation network. In: Proc. of the 23rd Int'l Conf. on Software Quality, Reliability, and Security. Chiang Mai: IEEE, 2023. 507–516. [doi: 10.1109/QRS60937.2023.00056]

[32] Zhao DD, Cheng WT, Zhou J, Wang HM, Li HH. Cancellable iris recognition scheme based on inversion fusion and local ranking. In: Proc. of the 30th Int'l Conf. on Neural Information Processing. Changsha: Springer, 2023. 340–356. [doi: 10.1007/978-981-99-8067-3_26]

[33] Liu XL, Wu CS, Menta M, Herranz L, Raducanu B, Bagdanov AD, Jui S, van de Weijer J. Generative feature replay for class-incremental learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020. 915–924. [doi: 10.1109/CVPRW50498.2020.00121]

[34] CASIA. CASIA iris image database. 2016. http://biometrics.idealtest.org/

[35] Othman N, Dorizzi B, Garcia-Salicetti S. OSIRIS: An open source iris recognition software. Pattern Recognition Letters, 2016, 82: 124–131. [doi: 10.1016/j.patrec.2015.09.002]

[36] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Proc. of the 18th Int'l Conf. Munich: Springer, 2015. 234–241. [doi: 10.1007/978-3-319-24574-4_28]

[37] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proc. of the 3rd Int'l Conf. on Learning Representations. 2015. 1–14.

[38] Gangwar A, Joshi A. DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In: Proc. of the 2016 IEEE Int'l Conf. on Image Processing. Phoenix: IEEE, 2016. 2301–2305. [doi: 10.1109/ICIP.2016.7532769]

[39] Chaudhry A, Dokania PK, Ajanthan T, Torr PHS. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In: Proc. of the 15th European Conf. on Computer Vision. Munich: Springer, 2018. 556–572. [doi: 10.1007/978-3-030-01252-6_33]

[40] Li ZZ, Hoiem D. Learning without forgetting. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018, 40(12): 2935–2947. [doi: 10.1109/TPAMI.2017.2773081]

[41] Rebuffi SA, Kolesnikov A, Sperl G, Lampert CH. iCaRL: Incremental classifier and representation learning. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5533–5542. [doi: 10.1109/CVPR.2017.587]

[42] Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, Hadsell R. Overcoming catastrophic forgetting in neural networks. Proc. of the National Academy of Sciences of the United States of America, 2017, 114(13): 3521–3526. [doi: 10.1073/pnas.1611835114]

[43] Douillard A, Cord M, Ollion C, Robert T, Valle E. PODNet: Pooled outputs distillation for small-tasks incremental learning. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 86–102. [doi: 10.1007/978-3-030-58565-5_6]

[44] Zhao BW, Xiao X, Gan GJ, Zhang B, Xia ST. Maintaining discrimination and fairness in class incremental learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2023. 13205–13214. [doi: 10.1109/CVPR42600.2020.01322]

[45] Yan SP, Xie JW, He XM. DER: Dynamically expandable representation for class incremental learning. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 3013–3022. [doi: 10.1109/CVPR46437.2021.00303]

[46] Wang FY, Zhou DW, Ye HJ, Zhan DC. FOSTER: Feature boosting and compression for class-incremental learning. In: Proc. of the 2022 European Conf. on Computer Vision. Tel Aviv: Springer, 2022. 398–414. [doi: 10.1007/978-3-031-19806-9_23]

[47] Petit G, Popescu A, Schindler H, Picard D, Delezoide B. FeTrIL: Feature translation for exemplar-free class-incremental learning. In: Proc. of the 2023 IEEE/CVF Winter Conf. on Applications of Computer Vision. Waikoloa: IEEE, 2023. 3900–3909. [doi: 10.1109/WACV56688.2023.00390]

[48] Zhou DW, Wang FY, Ye HJ, Zhan DC. PyCIL: A Python toolbox for class-incremental learning. Science China Information Sciences,

2023, 66(9): 197101. [doi: 10.1007/s11432-022-3600-y]

附中文参考文献:

[1] 宋继伟, 程多福, 王文峰. 生物特征识别技术与标准化. 北京: 中国电子技术标准化研究院, 2023. 42–53.

**赵冬冬**(1989－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为隐私保护, 虹膜识别, 数据挖掘.

**闫江**(2003－), 男, 本科生, 主要研究领域为隐私保护.

**宋宝刚**(2001－), 男, 硕士生, 主要研究领域为隐私保护, 虹膜识别, 数据挖掘.

**向剑文**(1975－), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为网络安全, 可信计算, 软件工程.

**廖虎成**(1996－), 男, 硕士生, 主要研究领域为隐私保护, 虹膜识别.