基于同态密文转换的隐私保护卷积神经网络推理方案

李瑞琪[1]，易琴[1]，黄艺璇[2]，贾春福[2,3]
（1. 中国民航大学安全科学与工程学院，天津 300300；2. 南开大学网络空间安全学院，天津 300350；
3. 天津市网络与数据安全技术重点实验室，天津 300350）

摘　要：为了解决现有隐私保护卷积神经网络交互频繁、推理准确率稍低等问题，基于同态密文转换框架，提出了一种同态友好型的非交互式隐私保护卷积神经网络推理方案。利用Pegasus同态密文转换框架，在卷积层中利用CKKS（Cheon-Kim-Kim-Song）密文进行并行化的卷积运算；在激活层和池化层中利用LWE密文和LUT（look-up table）技术实现激活函数、最大池化和全局池化的计算；利用Pegasus框架提供的密文转换技术，实现不同形式的同态密文之间的转换。理论分析和实验结果表明，所提方案能够保证数据安全，并且具有较高的推理准确率和较低的计算和通信开销。

关键词：隐私保护；卷积神经网络；同态加密；密文转换
中图分类号：TP309.2
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024216

Privacy-preserving convolutional neural network inference scheme based on homomorphic ciphertext transformation

LI Ruiqi[1], YI Qin[1], HUANG Yixuan[2], JIA Chunfu[2,3]
1. College of Safety Science and Technology and Engineering, Civil Aviation University of China, Tianjin 300300, China
2. College of Cyber Science, Nankai University, Tianjin 300350, China
3. Tianjin Key Laboratory of Network and Data Security Technology, Tianjin 300350, China

Abstract: To solve the problems of frequent interaction and low prediction accuracy of existing privacy-protected convolutional neural networks, a homomorphic friendly non-interactive privacy-protected convolutional neural network inference scheme was proposed via homomorphic ciphertext transformation. Utilizing the Pegasus framework, CKKS (Cheon-Kim-Kim-Song) ciphertext was used to parallelize convolution operations in convolution layer. In the activation layer and pooling layer, LWE ciphertext and LUT (look-up table) technology were used to calculate the activation function, maximum pooling and global pooling. Using the ciphertext conversion technology provided by the Pegasus framework, the conversion between different forms of homomorphic ciphertext is realized. Theoretical analysis and experimental results show that the proposed scheme can ensure data security, and has higher inference accuracy and lower calculation and communication overhead.

Keywords: privacy-preserving, convolutional neural network, homomorphic encryption, ciphertext transformation

0　引言

随着互联网的蓬勃发展，数据量呈指数级增长，大数据和云计算已成为引领信息技术领域发展的重要驱动力。在云计算时代，数据通常集中存储于云服务器端。随着云计算和大数据的结合，用户可以使用机器学习算法对海量、复杂的数据进行分

收稿日期：2024–10–08
通信作者：贾春福，cfjia@nankai.edu.cn
基金项目：天津市教委科研计划基金资助项目（No.2022KJ066）
Foundation Item：The Education Committee Research Program of Tianjin（No.2022KJ066）

析处理，这极大地提升了资源利用率和计算分析效率，可为个人、企业和机构提供更加精确、智能的决策支撑。在金融商业、医疗健康、城市管理以及科学研究等重要领域，机器学习外包计算都发挥着重要的推动作用。然而，数据的存储和处理服务往往都由第三方云服务提供商完成，随之而来的隐私泄露问题也日益凸显。

卷积神经网络（CNN, convolutional neural network）作为常见的深度学习模型，在图像分类等视觉识别任务中表现出色。在医疗诊断领域，可以使用卷积神经网络进行医学影像的分析，进行肿瘤、视网膜病变等疾病的辅助诊断；在安全监控领域，可以使用卷积神经网络进行人脸检测和识别，提高安全系统的性能；在电子商务领域，可以使用卷积神经网络识别商品图片，用于商品分类，提升搜索效率。

机器学习算法的训练、推理过程需要大量的用户数据作为支撑以保证其有效性和准确性，其中的用户隐私数据安全保护问题是大数据时代机器学习技术所面临的重大风险与挑战。当用户选择将数据外包至第三方云服务提供商进行外包计算时，需要保护隐私数据在传输、计算、分析等过程中的机密性、完整性和可用性，规避隐私数据泄露的风险。

由于个人隐私保护意识的不断增强和相关法律法规的出台，隐私数据的收集和使用受到了一定限制，这使传统的云计算环境下的机器学习的发展受到了阻碍和制约。隐私保护机器学习的出现和发展，在保持机器学习算法模型有效性和性能的同时，也保护了个人隐私数据的安全。

全同态加密（FHE, fully homomorphic encryption）[1-8]是常见的用于机器学习隐私保护场景的密码技术之一，其支持在密文上进行计算，能够在保证隐私数据的机密性的同时实现对密文数据的处理。将同态加密算法应用于机器学习的隐私保护时，隐私数据可以在加密状态下，被发送至第三方云服务器处进行机器学习模型的训练和推理，既节省了本地计算资源、提高了计算效率，又保护了隐私数据的安全。基于同态加密的机器学习隐私保护有交互式和非交互式2种类型。交互式隐私保护方案通常使用安全协议进行非多项式函数计算，其往往存在通信开销较高的问题；非交互式隐私保护方案使用逐位加密型同态加密方案或多项式近似方法进行非多项式函数计算，其面临计算开销过大或准确性较低的问题。因此，基于数据隐私保护和外包安全计算的需求，权衡计算开销、通信开销和准确性之间的关系，研究基于同态加密的非交互式机器学习算法隐私保护方案具有重大的现实意义。

1) 研究现状

深度学习模型在众多依赖隐私数据的领域（如金融、医疗保健等领域）有着出众的表现，因此深度学习中模型和数据的隐私保护得到了学界和工业界的广泛关注。隐私保护机器学习（PPML, privacy-preserving machine learning）领域近年来取得了很多成绩，发表了众多研究成果。安全推理旨在于不受信任的服务器中进行推理时保护用户隐私数据。安全推理由于其功能性和可行性得到了广泛研究。通过使用安全多方计算、安全第三方硬件（例如Intel SGX等）、全同态加密及其组合的技术，学者已经提出了各种安全推理方案[9-17]。其中，基于全同态加密的安全推理能够提供高安全性的数据隐私保护，支持在密文上进行计算，不需要客户与服务提供商之间进行频繁交互，基于全同态加密的安全推理方案是有发展前景的。

文献[9]首次基于全同态加密方案提出了密态神经网络隐私保护模型CryptoNets，该文使用全同态加密算法对数据集进行加密，通过将多个输入数据加密至一个密文提升计算效率，使用乘法深度极低的平方函数代替非线性的激活函数，在训练完成的卷积神经网络上实现加密数据安全推理，以加密形式向用户返回推理结果，在MNIST数据集上的推理准确率达到99%。但该模型仅适用于浅层神经网络，对于超过2个非线性层的神经网络，该方案推理准确率急剧下降。为解决这一问题，Chabanne[10]等保留平方函数作为激活函数这一思路，通过在每个非线性层前加入批量标准化层，使激活函数的输入具有稳定正态分布，提高卷积神经网络分类的准确性，但增加批量标准化层引入了额外的计算开销，增加了安全推理的时间。Chou[11]等提出了更快的方案，通过修剪神经网络参数来加速CryptoNets中的同态评估操作，减少同态乘法操作数量，使用较高次的多项式近似激活函数，提高推理准确性。文献[12]提出了低时延的神经网络安全推理模型，该文结合同态加密和混淆电路，设计了一种卷积的向量化表示，即单输入单输出（SISO,

single-input single-output）的同态卷积，可以对卷积层的同态密文输入进行更快的评估，且在非线性层通过两方计算协议进行激活函数的运算，该方案在MNIST数据集和CIFAR-10数据集上都有着良好的性能表现和较高的推理准确率。文献[13]提出使用TFHE方案来评估离散神经网络，虽然该方案可以支持在固定的开销下对任意深度的神经网络进行评估，且支持准确的激活函数评估，但由于单次同态操作的计算成本较高，带来较大的时间开销，且基于TFHE的方案[14-15]都需要对神经网络权重和输入进行离散化才能进行有效的评估。文献[16]结合同态加密和安全多方计算实现同态神经网络，基于同态加密实现线性层运算，基于安全多方计算，通过用户端与服务器端的交互计算实现非线性层运算，虽然实现了高准确率的安全推理，但引入了极高的通信开销。文献[17]提出了一个基于CKKS（Cheon-Kim-Kim-Song）同态加密方案的卷积神经网络推理方案，该方案使用CKKS技术加密图像像素点，增加一台云服务器并通过协议的方式来实现密文比较从而实现激活层和池化层。该方案虽然能够避免客户端与服务器的频繁通信，也能保证激活层和池化层的计算准确度，但是客户端加密的开销很大，并且也引入了额外的服务器。文献[18]利用混淆电路实现非线性层运算，并离线计算乘法三元组。该方案时间开销低，准确率高，同时避免客户端和服务器端的交互，但需要部署两台服务器。

基于上述文献的研究现状可知，在现有的基于同态加密的非交互式卷积神经网络隐私保护方案中，多数方案使用平方激活函数或多项式近似表达代替激活函数，这将对推理准确性造成影响；基于混合方案的交互式卷积神经网络隐私保护方案在使用安全协议进行激活函数计算时，会引入较大的通信开销。因此，基于上述研究现状，本文对基于同态加密的非交互式卷积神经网络隐私保护方案的研究具有必要性。

2) 本文贡献

本文针对目前机器学习隐私保护存在的问题和现有解决方案的不足，研究基于同态密文转换框架的卷积神经网络推理的隐私保护问题，以期实现用户可离线的密文上的非交互式神经网络推理方案，在不增加用户端额外的通信和计算开销并且也不需要增加额外的服务器的同时，能达到较高的推理准确率。本文基于Pegasus同态密文转换框架设计了非交互式卷积神经网络隐私保护方案，并给出了方案的性能、准确性和安全性分析。本文的工作主要包含以下几个方面。

① 采用CKKS全同态加密方案加密输入数据，通过CKKS支持的明文槽技术实现高吞吐率。

② 采用密文转换框架，针对卷积神经网络的不同运算，选用不同的加密算法以保证计算效率和准确性。在进入激活函数层前，将CKKS密文转换为LWE（learning with errors）密文，在LWE密文上通过LUT（look-up table）技术进行激活函数运算，不需要通过安全协议进行通信交互。

③ 使用LUT技术和最大树算法实现最大池化操作，相比平均池化层，最大池化层能提升模型分类效果；使用同态友好的全局平均池化层代替全连接层，减少模型参数，提高同态计算效率。

④ 实现卷积神经网络隐私保护框架，对所提方案性能、准确性和安全性进行分析。与交互式隐私保护方案相比，所提方案极大地降低了通信开销和计算开销，与非交互式隐私保护方案相比，所提方案具有更高的准确率。此外，所提方案还具有可拓展性，适用于任意深度的卷积神经网络。

1 基础知识

本节将介绍所提方案需要的基础知识，包括Pegasus同态密文转换框架、卷积神经网络以及方案系统模型。

首先对本文中使用的符号进行说明。本文使用粗体小写字母表示向量，如$\boldsymbol{a}$；使用a_j表示向量$\boldsymbol{a}$的第j个分量；使用$\boldsymbol{a} \ll k$表示向量分量向左旋转k位。使用粗体大写字母表示矩阵，如$\boldsymbol{M}$；使用$\boldsymbol{M}_{i,j}$表示矩阵$\boldsymbol{M}$中第i行第j列的元素。使用$\{a_i\}_{1 \leqslant i \leqslant n}$表示集合$\{a_1,a_2,\cdots,a_n\}$。

1.1 Pegasus同态密文转换框架

文献[19]中提出了一种名为Pegasus的同态密文转换框架，其支持在CKKS密文和LWE密文之间相互转换，从而可以在一个隐私保护方案中实现在CKKS密文上进行实数域上的高效且并行化的多项式运算，也可在LWE密文上进行非多项式运算的LUT运算。

设ct为CKKS密文，其加密的多项式为$\hat{m}$，$\hat{m}$由向量$\boldsymbol{v}$编码得到的。在Pegasus密文转换框架中可以进行以下操作。

1) Rotation 算法。输入 CKKS 密文 ct、正整数 k；输出密文 $\mathrm{ct}' = \mathrm{Rotate}(\mathrm{ct},k)$，其对应的明文多项式解码后得到的向量为 $\boldsymbol{v} \ll k$。

2) Coeff-Extract 算法。输入 CKKS 密文 ct，正整数 k；输出一个的 LWE 密文 $c = \mathrm{Extract}(\mathrm{ct},k)$，其加密了 $\hat{m}$ 的第 k 个系数。

3) Slot-to-Coeff 算法。输入 CKKS 密文 ct；输出一个新的 CKKS 密文 $\mathrm{ct}' = \mathrm{SlotToCoeff}(\mathrm{ct})$，其加密了多项式 $\hat{v}$，且 $\hat{v} = \sum_{i=0}^{N-1} v_i X^i$。

4) LUT 算法。输入一个加密了 m 的 LWE 密文 c 和函数 $T(x)$；输出一个新的 LWE 密文 $c' = \mathrm{LUT}(c,T(x))$，其对应的明文为 $T(m)$。

5) Repack 算法。输入一组 LWE 密文 $\{c_i\}_{1\leqslant i\leqslant k}$，其中 c_i 对应的明文为 m_i；输出一个 RLWE 密文 $\mathrm{ct} = \mathrm{Repack}(c_1,c_2,\cdots,c_k)$，其对应的明文多项式解码后得到的向量为 $\boldsymbol{m} = (m_1,m_2,\cdots,m_k)$。

密文转换框架结合了不同同态加密方案在多项式运算和非多项式运算上的优势，在进行实际应用时，通常根据运算特性的需求，转换至合适的同态密文形式，以提高计算的准确性和效率。机器学习算法通常包含大量多项式运算和非多项式运算，将密文转换框架应用于机器学习隐私保护方案中可以有效地提高计算效率和模型准确率。

1.2　卷积神经网络

CNN 是一种深度的前馈神经网络模型，主要用于解决图像分析处理问题。CNN 的核心思想是通过多层网状结构进行特征映射、池化、局部连接等操作以模仿生物视觉感知的过程。近年来，CNN 在语音识别、图像分类、目标检测、医学诊疗等领域都有着突破性进展。

经典的 LeNet-5 模型[20]如图 1 所示，其基本结构主要由以下几个部分组成。

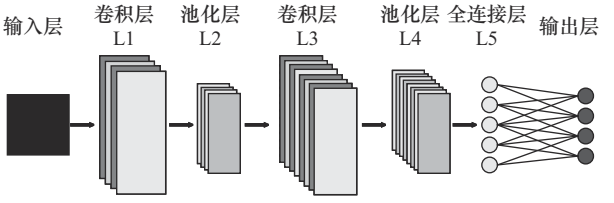

图 1　经典的 LeNet-5 模型

卷积层。CNN 模型中的核心层，主要进行大量的卷积运算，即图像和卷积核对应区域做内积运算。卷积核是一个权值矩阵，用于表示一个像素与周围像素之间的联系，每个卷积核都是一种特征提取器，进行运算后可以生成新的特征图，卷积层的运算结果决定着多层神经网络的分类效果。卷积核按照步长不断平移进行局部感知，与感受野内的数据进行卷积运算，生成新的二维特征激活图，提取出隐藏于图像数据内部的特征结构。

池化层。数据在经过卷积层提取特征得到特征图后，通过池化层降低数据的空间维度，从而减少后续计算的成本，并且有助于控制过拟合。最常用的池化方法是最大池化，它选择每个区域的最大值作为输出，将各区域的最大值重新连接形成一个新的特征图。池化操作能维持图像特征不变的同时使得特征图的维数显著下降。

全连接层。所有神经元和上一层的每个神经元都存在权重连接。全连接层将经过卷积、池化操作提取的局部特征整合为目标特征类数量的输出，即将局部特征映射至样本标记空间中。

传统方案在分类层的选取中通常选用全连接层，一般需要 2 个紧密连接的全连接层实现分类。然而，全连接层的权重参数在神经网络模型参数中占极高比例，会引入较大的计算开销。因此，所提方案使用全局平均池化层作为分类层，所提方案使用的 CNN 模型如图 2 所示。

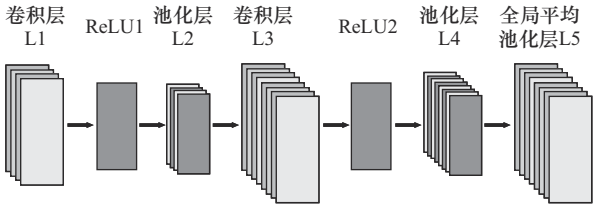

图 2　所提方案使用的 CNN 模型

全局平均池化层。选用和输入图像相同尺寸的池化窗口，对整体输入数据进行取平均值运算，输出大小为 1×1 的数据。全局平均池化操作极大地减少了分类层中的权重参数数量，可以有效提高同态卷积神经网络效率。

1.3　系统模型

所提方案提出的卷积神经网络隐私保护框架的主体有两部分，用户端和服务器端。所提方案的整体框架如图 3 所示，用户持有待预测的隐私数据；服务器拥有训练良好的卷积神经网络模型，该模型可为用户提供图像推理、图像分类等服务。同时，

为完成密文数据上的推理任务，模型中的参数也需相应地进行同态加密方案中环上多项式的编码，以便后续的同态运算。

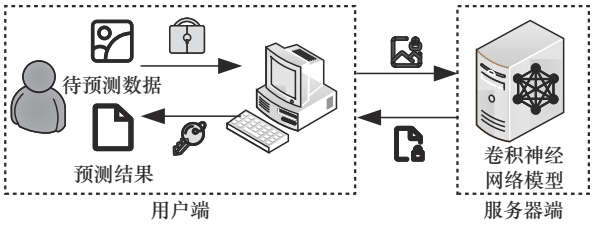

图3　所提方案的整体框架

假设计算服务器是诚实且好奇的，会按照方案进行计算，但也不可避免地会尝试获取原始数据和预测结果。为避免隐私数据泄露，用户端的数据所有者选择将数据经CKKS全同态加密方案加密后上传至服务器端，服务器进行卷积神经网络推理计算，将密文推理结果返回给用户端，用户端对其解密得到最终预测结果。基于全同态加密的特性，即支持不需要解密的情况下在密文上进行卷积神经网络推理计算，其能够用于实现安全高效的深度学习模型外包计算。具体流程如下。

1) 用户将待预测的本地数据进行同态加密并将密文发送给计算服务器。

2) 计算服务器接收数据，按卷积神经网络预测方案顺序，依次将上一步结果输入卷积层、激活层、池化层、全局平均池化层，最后得到最终的预测结果密文，并将其返回给用户。

3) 用户使用私钥解密预测结果密文，得到最终预测结果。

所提方案中的卷积神经网络隐私保护框架设计目标如下。

1) 安全性。保护用户端数据所有者的隐私数据安全和服务器端卷积神经网络模型安全。数据所有者通过同态加密方式对数据进行隐私保护，上传至服务器端进行卷积神经网络的推理计算。在安全推理阶段，数据始终保持加密状态，服务器模型的参数也为密文形式。服务器无法获取任何数据的相关信息。这能保证密文数据的机密性和模型安全。

2) 准确性。服务器提供高准确性的同态卷积神经网络推理服务。所提方案采用的是支持在CKKS密文和LWE密文之间相互转换的同态密文转换框架，在线性层使用CKKS加密方案，其支持在密文上进行高效同态卷积运算；在非线性层使用LWE加密方案，其支持在密文上进行激活函数运算和最大池化操作。以此保证安全推理的准确性。

3) 高效性。保证外包计算过程的高效性。从用户上传数据开始，服务器进行同态卷积神经网络推理过程，最后返回推理结果至用户端，整个过程所需的运行时间合理。另外，所提方案除数据上传和返回结果阶段，在服务器端进行卷积神经网络安全推理时，不需要与用户端进行额外通信交互，允许在用户端离线状态下进行服务器安全推理操作，保证卷积神经网络安全推理的高效性。

为实现安全高效的同态卷积神经网络推理计算，本文对卷积神经网络结构进行优化，设计了安全、准确、高效的同态卷积神经网络模型。针对卷积，所提方案使用同态卷积运算方法实现高吞吐的卷积运算，降低复杂全同态加密方案带来的计算开销；针对激活层，所提方案使用密文转换框架将CKKS密文转换为LWE密文以进行激活函数的LUT计算，而非使用多项式近似激活函数，影响推理准确性；针对池化，转换后的LWE密文支持最大值计算，可以在较浅层卷积神经网络中采用最大池化操作，获得较为显著的图像特征分类结果；在进入下一个卷积层时，通过密文转换框架中的重打包技术，将LWE密文转换回CKKS密文，再次进行高效并行的卷积运算；最后，使用全局平均池化层代替全连接层，减少模型参数，降低同态复杂计算的开销。

2　具体方案

本节将具体介绍基于同态密文转换框架的隐私保护卷积神经网络推理方案，具体包括：预处理算法、卷积层算法、激活层算法、最大池化层算法、全局平均池化层算法。

2.1　预处理算法

图像数据在加密前，先进行归一化和缩放操作，保证模型的性能和准确度。用户端使用CKKS加密方案和明文槽技术，将一张图像的每个像素加密至一个CKKS密文中，可以有效地提高后续计算效率。输入图像的尺寸为$1 \times w \times w$，在对输入图片进行归一化的预处理后，按照行列的顺序将其加密至一个CKKS密文中，如图4所示。

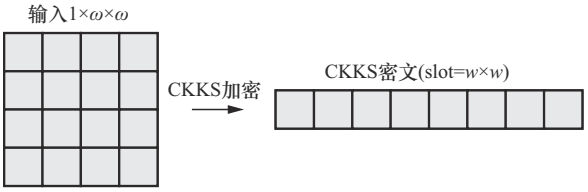

图4　输入图像的加密方式

服务器端中训练良好的CNN模型中参数也需编码至同态加密中环上多项式的形式，以便于进行后续推理阶段中的计算。

2.2　卷积层算法Convolution

所提方案使用在卷积神经网络中广泛使用的2D卷积，并在全同态密文上进行2D卷积的同态计算。

设输入为一个二维数组 $\boldsymbol{I} \in \mathbb{R}^{w \times w}$，卷积核 $\boldsymbol{K} \in \mathbb{R}^{k \times k}$，单个2D卷积 $\mathrm{Conv}(\boldsymbol{I},\boldsymbol{K}) \in \mathbb{R}^{d \times d}$（填充采用valid padding方式）定义如下

$$\mathrm{Conv}(\boldsymbol{I},\boldsymbol{K})_{i,j} = \sum_{i'=0,j'=0}^{k} K_{i',j'} I_{i+i',j+j'}$$

其中，i,j,w,k,d的关系为$0 \leqslant i,j < d = w - k + 1$。

所提方案使用文献[12]提出的同态卷积算法加速卷积层运算。输入$\boldsymbol{I} \in \mathbb{R}^{w \times w}$经同态加密后得到一个同态密文。首先将该同态密文进行多次旋转以便于与卷积核对齐，然后将$k \times k$个旋转密文与相对应的卷积核$\boldsymbol{K} \in \mathbb{R}^{k \times k}$进行同态乘法，将所有乘法结果相加后，即得到包含卷积结果的同态密文；最后使用明文掩码删除不相关的明文槽，即可得到最终卷积结果。因此，同态卷积操作需要进行$k^2 - 1$次同态旋转和k^2次同态乘法操作。以3×3的输入数据和2×2的卷积核为例，具体过程如图5所示。输入密文经三次旋转后，原始密文和得到的旋转密文都和对应的卷积核参数进行乘法运算，最后将4个乘法结果相加，得到一个包含最终卷积结果的密文，图中最下方的向量中包含的即卷积结果$C_{00},C_{01},C_{10},C_{11}$。

所提方案的卷积层算法如算法1所示，对输入的CKKS密文进行多次旋转，将旋转密文和对应的卷积核参数相乘即可快速计算出同态卷积。数据在经卷积运算后，将在下一步进行激活函数运算。

算法1　卷积层算法

输入　CKKS密文C^{CKKS}，循环移位位数l，卷积核$\boldsymbol{K} = (K_{0,0},K_{0,1},\cdots,K_{k-1,k-1})$

输出　CKKS密文卷积结果$C_{\mathrm{conv}}^{\mathrm{CKKS}}$

1) for $i = 0$ to $k^2 - 1$

2)　$A_i = \mathrm{Rotate}(C^{\mathrm{CKKS}},l)$;

3) end for

4) for $i = 0$ to $k - 1$;

5)　for $j = 0$ to $k - 1$;

6)　　$\mathrm{Tmp}_i = A_i K_{ij}$;

7)　　$\mathrm{Sum} = \mathrm{Sum} + \mathrm{Tmp}_i$;

8)　end for

9) end for

10) 返回$C_{\mathrm{conv}}^{\mathrm{CKKS}} = \mathrm{Sum}$。

2.3　激活层算法

激活函数可以引入非线性因素，提高神经网络模型对于非线性的表达能力，学习数据中线性不可分的复杂模式。在卷积神经网络模型中，常用的激活函数有Sigmoid、Tanh、ReLU函数，这些函数都不是可以用多项式表示的函数，而CKKS同态加密方案不支持直接进行非多项式函数的运算。因此，所提方案利用Pegasus同态密文转换框架，将CKKS密文转换为LWE密文，在LWE密文上进行LUT运算来实现激活函数。

激活层算法如算法2所示，其中的$T(x)$可以选用任意激活函数，length为卷积层计算结果的元素个数。该算法首先将输入的一个CKKS密文通过Slot-To-Coeff算法和Coeff-Extract算法转换为多个LWE密文，然后在LWE密文上准确计算激活函数，最后输出LWE密文的计算结果。数据在完成激活函数运算后，

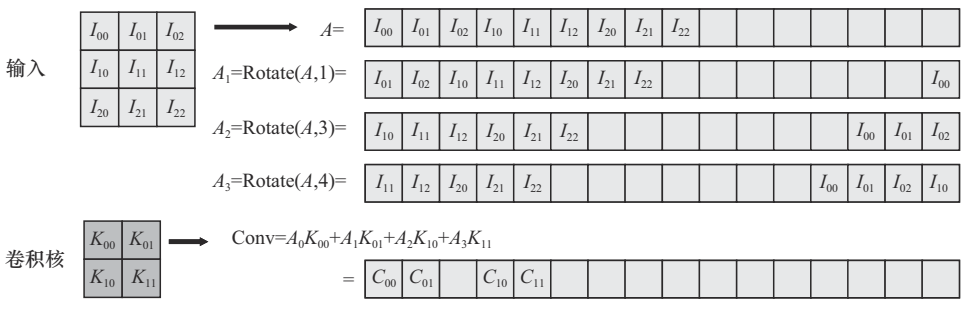

图5　卷积层计算过程

将以LWE密文进入池化层，进行最大池化操作。

算法2　激活层算法

输入　卷积结果CKKS密文$C_{\text{conv}}^{\text{CKKS}}$

输出　激活函数运算结果LWE密文$\{C_0^{\text{FHEW}},C_1^{\text{FHEW}}\cdots,C_{\text{length}-1}^{\text{FHEW}}\}$

1) $C_{\text{tmp}}=\text{SlotToCoeff}(C_{\text{conv}}^{\text{CKKS}})$；

2) $\{\text{Tmp}_i\}_{0<i<\text{length}}=\text{ExtractCoeff}(C_{\text{tmp}})$；

3) for $i=0$ to length -1；

4)　$C_i^{\text{FHEW}}=\text{LUT}(T(x),\text{Tmp}_i)$；

5) end for；

6) 返回$\{\text{C}_{\text{i}}^{\text{FHEW}}\}_{0\leqslant i<\text{length}}$。

除了需要解决如何进行同态激活函数运算这个问题之外，还需解决的一个问题是卷积层与激活层的连接，如图6所示。

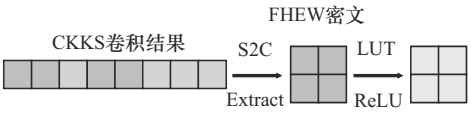

图6　卷积层与激活层的连接

在进行同态激活函数运算前，使用密文转换框架中的Slot-to-Coeff算法和Coeff-Extract算法，将卷积层的有效输出从卷积结果密文中提取出来并将其转换为LWE密文，再在LWE密文上使用LUT函数对卷积层的输出执行激活函数运算。

2.4　最大池化层算法

在此前的方案中，由于CKKS方案等RLWE型FHE方案不支持在非交互的情况下实现密文大小比较操作，因此在同态卷积神经网络隐私保护方案中，通常在池化层中使用对同态加密更加友好的平均池化操作来代替最大池化操作。然而，文献[21]已经证实最大池化操作在众多图像分类任务中表现更为突出；而LWE密文是支持密文大小比较操作的，因此所提方案选择进行最大池化，最大池化过程如图7所示。

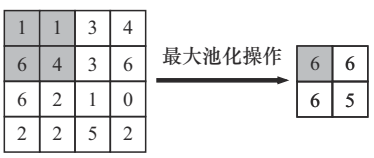

图7　最大池化过程

所提方案通过进行密文转换，将CKKS密文转换为LWE密文以进行密文大小比较运算。下面本文将参考文献[19]中的最大值计算方法，介绍在同态密文上实现同态比较运算的过程。

已知以下运算可输出2个密文之间的最大值：对于任意两数m_0和m_1，它们的最大值可以表示为

$$\max(m_0,m_1)=0.5(m_0+m_1)+0.5|m_0-m_1|$$

在LWE密文上，利用LUT函数$T_0(x)=0.5x$和$T_1(x)=0.5|x|$即可输出2个同态密文之间的最大值。而在一组密文之间运算输出其中的最大值可以使用最大树算法，其流程如图8所示。

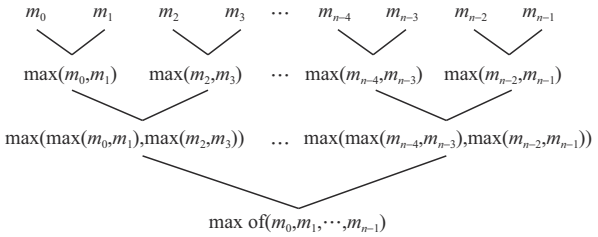

图8　最大树算法流程

计算输出t个加密元素的最大值通常需要进行$O(2t)$次的LUT操作。但在最大池化操作中，由于t值一般很小（例如$t=4$），可以将最大树中求最大值的算法修改为

$$\max'(m_0,m_1)=m_0+m_1+|m_0-m_1|$$

修改后的算法输出的最大值变成了原最大值的t倍，但是倍数t可以通过一个LUT函数$T(x)=\dfrac{x}{t}$去除，所以$T(x)$的输出即一组密文中的最大值。因此，当t值较小的情况下，可以通过使用$O(1)$次LUT操作来实现最大池化操作，极大地提升了池化操作的效率。最大池化层算法如算法3所示，算法输入和输出均为LWE形式密文。

算法3　最大池化层算法

输入　$a\times a$个LWE密文激活函数运算结果$\{C_{i,j}^{\text{FHEW}}\}_{0\leqslant i<a,0\leqslant j<a}$，池化核大小$p\times p$，$T(x)=\dfrac{x}{p^2}$

输出　最大池化运算结果的LWE密文$\{C_{\text{Max},i}^{\text{FHEW}}\}_{0\leqslant i<\left(\frac{a}{p}\right)^2}$

1) for $i=0$ to $\dfrac{a}{p}-1$；

2)　for $j=0$ to $\dfrac{a}{p}-1$；

3)　$P_{ij}=\max(C_{pi,pj}^{\text{FHEW}},\cdots,C_{pi,pj+p-1}^{\text{FHEW}},C_{pi+1,pj}^{\text{FHEW}},\cdots,$

$C^{\mathrm{FHEW}}_{pi+1,pj+p-1},\cdots,C^{\mathrm{FHEW}}_{pi+p-1,pj},\cdots,C^{\mathrm{FHEW}}_{pi+p-1,pj+p-1}$);

4)　end for；

5) end for；

6) for $i = 0$ to $\frac{a}{p} - 1$；

7)　for $j = 0$ to $\frac{a}{p} - 1$；

8)　$C^{\mathrm{FHEW}}_{\mathrm{Max},pi+j} = \mathrm{LUT}(T(x), P_{ij})$；

9)　end for；

10) end for；

11) 返回 $\{C^{\mathrm{FHEW}}_{\mathrm{Max},i}\}_{0 \leqslant i < \left(\frac{a}{p}\right)^2}$

2.5　全局平均池化层算法 Averagepooling

传统方案在分类层的选取中通常选用全连接层，一般需要 2 个紧密连接的全连接层实现分类。全连接层的每个神经元都与上一层的每个神经元连接。全连接层综合上一层的输出特征，并在下一层输出分类结果。但是，全连接层的权重参数在神经网络模型参数中占极高比例，会引入较大的计算开销。因此，所提方案使用全局平均池化层作为分类层，全局平均池化过程如图 9 所示。

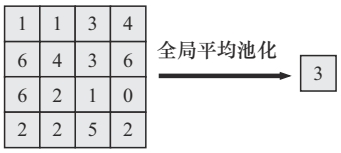

图 9　全局平均池化过程

全局平均池化算法如算法 4 所示，该算法首先计算所有输入数据的和值，然后使用 LUT 函数对和值进行除法运算以取所有元素的平均值，最后输出全局平均池化结果。

算法 4　全局平均池化层算法

输入　$g \times g$ 个 LWE 密文 $\{C^{\mathrm{FHEW}}_i\}_{0 \leqslant i < g^2}$，LUT 函数 $T(x) = \frac{x}{g^2}$

输出　全局平均池化运算结果 LWE 密文 $C^{\mathrm{FHEW}}_{\mathrm{avr}}$

1) for $i = 0$ to $g^2 - 1$；

2)　Sum = Sum + C^{FHEW}_i；

3) end for；

4) $C^{\mathrm{FHEW}}_{\mathrm{avr}} = \mathrm{LUT}(T(x), \mathrm{Sum})$；

5) 返回 $C^{\mathrm{FHEW}}_{\mathrm{avr}}$。

2.6　最大池化层与卷积层的连接

在上述算法的基础上，实现完整的隐私保护神经网络推理方案仍需解决一个问题：第一次最大池化层的输出如何转化为第二次卷积层的输入。本文采用密文转换的方式，灵活地在 CKKS 密文和 LWE 密文间切换，在 CKKS 密文上进行卷积层、全连接层中大量的同态乘法运算，在 LWE 密文上进行大小比较、激活函数等非多项式运算，不需要在卷积神经网络的安全推理阶段中与用户进行额外的通信交互。在所提方案的卷积神经网络隐私保护框架中，密文形式转换情况如图 10 所示。

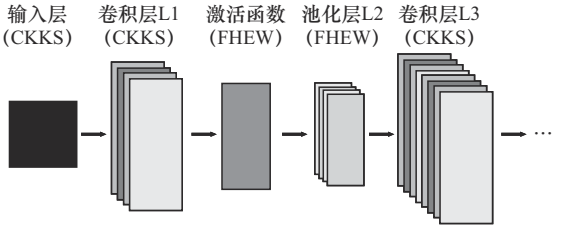

图 10　所提方案中密文形式转换情况

当 CKKS 密文在卷积层进行卷积计算后，首先通过 Slot-To-Coefficients 算法和 Coeff-Extract 算法将 CKKS 密文转换为 LWE 密文，然后在 LWE 密文上执行激活函数的 LUT 计算，得到准确的激活函数计算结果。在池化层中，在 LWE 密文上执行最大树算法输出池化窗口内加密数据的最大值，完成最大池化操作。在进入下一个卷积层前，调用 Repack 算法，将多个 LWE 密文重打包至一个 CKKS 密文中，以进行后续的高效同态卷积运算。

3　方案分析

3.1　安全性分析

所提方案中的安全模型参考文献[22]中的隐私保护机器学习方案的安全模型，服务器是诚实且好奇的，拥有卷积神经网络模型，用户端是数据所有者，安全目标是防止不受信任的服务器访问用户的原始隐私数据。用户端在使用全同态加密方案进行数据加密后，将隐私数据发送至不受信任的服务器。服务器直接对加密数据进行机器学习模型推理，在推理阶段不涉及数据解密，也不需要与用户端进行数据交互，最后返回推理结果的密文至用户端。只有持有私钥的用户端才能解密最终推理结果，保证数据的隐私安全。

所提方案的安全目标为云服务器和具备窃听能力的敌手无法获得用户隐私数据，也无法通过同态卷积神经网络模型推断原始数据的信息。所提方案

的安全性可以通过如下定理来说明。

定理1 在所提方案中，云服务器和具备窃听能力的敌手依据其获取的信息，成功获得用户隐私数据相关信息的概率是可以忽略的。

证明 云服务器或具备窃听能力的敌手可以通过窃听信道获取上传到云服务器的用户图像数据的密文。用户将图像数据经CKKS加密算法加密上传。数据在进行安全推理时，始终保持CKKS同态密文形式和LWE同态密文形式。已知CKKS和LWE这2个同态加密方案是具有CPA安全性的，因此云服务器或具备窃听能力的敌手成功恢复用户原始敏感数据的概率是可忽略的。在系统实现中，若选取同态加密标准化组织设定的安全级别和具体参数，那么攻破CKKS加密算法和LWE加密算法这2个加密算法的概率是可以忽略的。同样地，若敌手获取到同态卷积神经网络推理阶段的密文，则在此过程中成功恢复用户原始敏感数据的概率仍然等同于攻破CKKS加密算法和LWE加密算法的概率，这一概率是可以忽略的。证毕。

3.2 计算开销分析与对比

本节对所提方案各层的计算开销进行理论上的分析并与文献[17]方案进行对比。设输入图像的尺寸为$1 \times w \times w$；卷积核的尺寸为$k \times k$；输出通道为1；最大池化窗口的尺寸为$p \times p$；最大池化层的输出尺寸为$g \times g$。

在预处理过程中，可以设定CKKS密文的明文槽数大于或等于$1 \times w \times w$，那么就可以将一张乃至多张图片同时加密到一个CKKS密文中。

在同态卷积层中，使用卷积层算法进行同态卷积操作所需的同态加法次数为k^2-1，同态乘法次数为k^2，循环移位次数为k^2-1。卷积层的输出仍为一个密文，其加密了尺寸为$(w-k+1)^2$的矩阵。

在同态激活层中，需要使用一次Slot-to-Coeff算法以及$(w-k+1)^2$次Coeff-Extract算法将卷积层输出的CKKS密文转换为$(w-k+1)^2$个LWE密文，然后在这些LWE密文上使用LUT算法，而Slot-to-Coeff算法和Coeff-Extract算法开销较小。因此，激活层的主要开销为进行$(w-k+1)^2$次LUT运算。

在最大池化层中，在池化窗口范围内使用改进最大树算法，输出池化窗口内的最大值作为池化层输出结果，池化层输出图像的尺寸为$\left[\frac{w-k+1}{p}\right]^2$，因而需进行$\left[\frac{w-k+1}{p}\right]^2(p^2+1)$次LUT运算。在进入下一个卷积层前，需要执行Repack算法，将LWE密文重新转换为CKKS密文以便于进行同态卷积运算。

在全局平均池化层中，将密文图像矩阵的每个元素相加后，将结果与函数$T(x)=\frac{x}{g^2}$输入LUT算法，即可输出密文图像矩阵的所有元素的平均值。在同态全局平均池化层中，需要进行$g \times g$次同态加法和一次LUT运算。

文献[17]方案将一个像素加密至一个CKKS密文中，在卷积层需要做$[(w-k+1)]^2k^2$次同态乘法和$(w-k+1)^2(k^2-1)$次同态加法，虽然不用做旋转操作，但需要进行大量的同态加法和同态乘法，且次数都与图片尺寸大小相关，开销较大。激活层通过解密成明文后与另一台服务器交互来实现，因而不需要同态操作。池化层同样基于明文上的操作进行，然后再次加密并做$\left[\frac{w-k+1}{p}\right]^2$次同态乘法；设全连接层有$m$个神经元，需要做$m \times g \times g$次同态乘法和$m \times g \times g$次同态加法操作。表1为所提方案与文献[17]方案的计算开销对比。

表1 与文献[17]方案计算开销对比

方案	同态乘法	同态加法	旋转	LUT
文献[17]方案	$O(k^2(w-k)^2)$	$O(k^2(w-k)^2)$	—	—
所提方案	$O(k^2)$	$O(k^2)$	$O(k^2)$	$O((w-k)^2)$

3.3 通信开销分析与对比

本节对所提方案在上传阶段、推理阶段和下载阶段的通信开销进行理论分析并与文献[17]方案进行对比。设输入图像的尺寸为$1 \times w \times w$；卷积核的尺寸为$k \times k$；输出通道为1；最大池化窗口的尺寸为$p \times p$；全连接层中的神经元个数为n。一个CKKS密文的尺寸记为l_{CKKS}，一个LWE密文的尺寸记为l_{LWE}。

所提方案采用CKKS的明文槽技术，将输入图片多个像素加密至一个CKKS密文上，因此上传阶段的通信开销为一个CKKS密文的大小；在推理阶段所提方案利用Pegasus密文转换框架实现了不需

要进行交互的激活函数计算和最大池化操作，仅在一台服务器上完成卷积神经网络推理全过程，因此所提方案在推理阶段没有通信开销；由于所提方案的全局平均池化层的输出为一个LWE密文（假设输出通道为1），因此所提方案在下载阶段的通信开销为一个LWE密文的大小。

文献[17]方案将一个像素加密成一个CKKS密文，一张$w \times w$的图像需要w^2个CKKS密文；而在实现激活层时，服务器1向服务器2发送卷积后的结果也就是$(w-k+1)^2$个密文，在云服务器2上实现激活层、池化层算法，然后向服务器1发送$\left[\frac{w-k+1}{p}\right]^2$个CKKS密文；服务器1实现全连接层，并向用户发送预测结果，为n个CKKS密文。表2展现的是与采用两台服务器实现的交互式隐私保护文献[17]方案预测一张图像通信开销对比情况。

表2　与文献[17]方案的各阶段通信开销对比

方案	上传阶段	推理阶段	下载阶段
文献[17]方案	$w^2 l_{\mathrm{CKKS}}$	$O((w-k)^2) l_{\mathrm{CKKS}}$	$n l_{\mathrm{CKKS}}$
所提方案	l_{CKKS}	0	l_{LWE}

3.4　功能对比分析

本节将所提方案与现有隐私保护方案进行比较，将从用户数据隐私性、方案准确性、交互性，以及是否需部署额外的云服务器等方面进行比较分析。

表3所列出的5个方案中均使用全同态加密算法，在用户私钥不泄露的情况下，恶意攻击者无法得到用户隐私数据。

表3　现有方案功能对比

方案	隐私性	准确性	非交互	单节点
CryptonNets[9]	√	×	√	√
SecureML[23]	√	√	×	√
CNNP[18]	√	√	√	×
文献[17]方案	√	√	√	×
所提方案	√	√	√	√

在模型准确性方面，CryptoNets[9]使用全同态加密算法对数据集进行加密，使用乘法深度极低的平方函数代替非线性的激活函数，但该模型仅适用于浅层神经网络，对于超过2个非线性层的神经网络，该方案推理准确率急剧下降。所提方案基于密文计算，且在非线性层使用密文转换，使用LWE密文进行非线性运算，能较准确地实现非线性层，因此所提方案在模型精确度无较大损失。

在用户和服务器的交互复杂度方面，SecureML[23]需要用户保持在线，须持续地和服务器进行交互，增加了用户的通信开销。所提方案中，用户仅需将加密后的数据上传至服务器即可离线等待最后结果，减少了用户通信开销。

部署云服务器的数量也应作为参考的维度，CNPP[18]和文献[17]方案使用均部署了额外的云服务器实现非激活层，虽然减少用户和服务器之间通信开销，但造成了资源消耗的增加，同时，更多的服务器意味着更多的潜在漏洞，增加数据泄露的风险。所提方案所提出的模型仅需要一台云服务器即可完成推理全过程。

综上，所提方案能够保证数据安全，实现卷积神经网络过程，且在非线性层利用密文转换在保证准确率较高的同时，用户和服务器无交互，并且具有较高的推理准确率和较低的计算和通信开销。

4　实验分析

本文进行明文神经网络训练和密文神经网络推理过程的配置如下，明文神经网络训练环境部署在win10操作系统上，处理器为Intel(R) Core(TM) i7-7200U @2.50 GHz，内存为16 GB，编程语言为Python3.6，工具库为Pytorch[64]。密文同态神经网络的推理环境部署在Ubuntu18.04.5操作系统上，处理器为Intel(R) Xeon(TM) Gold 6132 @2.60 GHz，内存为64 GB，编程语言为C++，工具库为OpenPegasus。

所提方案采用的同态密文转换框架中的加密方案所选取参数如下：在CKKS方案中，多项式次数为65 536，密文模数位数为239 bit，密文最大级数为5，调节因子为2^{40}，明文槽数为1 024；LWE方案中向量维数为1 024；LUT算法中选择的多项式次数为4 096。2个方案以及密文转换过程中选取的参数设定满足同态加密标准化组织设定的安全级别。

MNIST数据集是机器学习在图像分类领域中最经典的数据集之一，共有10类黑白手写数字图片，其中训练集包含60 000张图片，测试集包含10 000张图片，图片尺寸均为$1 \times 28 \times 28$。在HE-CNN的评估中，本文在明文上使用训练集中的数据训练网络模型参数，在密文上对测试集进行同态HE-CNN安

全推理评估。MNIST数据集是一个较简单的数据集，但在同态安全推理任务中，它是一个标准基准数据集。Fashion MNIST数据集是一个较MNIST数据集更接近普适物体的灰度图像服装图像数据集，共有10类黑白图片，其中训练集包含50 000张图片，测试集包含10 000张图片，图片尺寸均为1 × 28 × 28。Fashion MNIST数据集上的分类问题较MNIST数据集更为困难。在同态安全推理任务中，如何在更复杂的数据集实现较高的分类准确率成为研究热门。

用于MNIST数据集安全推理的网络架构及参数如表4所示，该网络使用ReLU激活函数和最大池化层，能够满足高准确性推理的需求，并引入了全局平均池化层，极大地减少了同态操作数，其为一个高准确性的同态友好型卷积神经网络。

表4 MNIST上同态卷积神经网络参数

网络层号	网络层类型	网络层具体参数	输入尺寸	输出尺寸
1	卷积层	卷积核大小为5×5	1×28×28	5×24×24
2	ReLU激活层	—	5×24×24	5×24×24
3	最大池化层	池化窗口为2×2	5×24×24	5×12×12
4	卷积层	卷积核大小为5×5	5×12×12	10×8×8
5	ReLU激活层	—	10×8×8	10×8×8
6	最大池化层	池化窗口为2×2	10×8×8	10×4×4
7	全局平均池化层	池化窗口为4×4	10×4×4	10×1×1

用于Fashion MNIST数据集安全推理的网络架构则比MNIST数据集上的网络结构多一层卷积层和最大池化层，以实现更复杂的分类任务。在硬件条件满足的情况下，可以在更深层的网络中使用所提方案的设计思想，达到更高的分类准确率。

4.1 推理准确率分析

如表5所示，在浅层卷积神经网络中，在MNIST数据集上进行安全推理时，所提方案的准确率为99.50%，而文献[17]方案的准确率为89.03%。

表5 与文献[17]方案的推理准确率对比

方案	准确率
文献[17]方案	89.03%
所提方案	99.50%

所提方案还使用更多层的网络结构对Fashion MNIST数据集进行安全推理，达到了90.0%的准确率。在理想状况下，所提方案在密文上推理的准确性与在明文上推理的准确性接近，能达到较高的准确性。所提方案实现了较高的安全推理准确率，使用密文转换框架，可以实现准确的激活函数计算，且具有较强的可扩展性，适用于任意深度的卷积神经网络模型，理论上在更深层的同态卷积神经网络安全推理计算中可以达到比先前方案更高的准确率。

4.2 计算开销分析

表6为所提方案在卷积神经网络上进行MNIST数据集的安全推理的各层计算开销。其中，卷积层的运算由于在CKKS密文上进行多项式运算，开销极小。由于激活函数、池化层和全局平均池化层涉及较多在LWE密文上进行的非多项式运算，这些层中使用了较多的LUT运算。安全推理的90%以上计算开销都由LUT运算产生，通过使用性能更加优秀的硬件、多核处理器的多线程计算以及后续对LUT运算加速的研究，所提方案的计算开销可以得到有效降低。

表6 所提方案计算开销

网络层	时间/s
卷积层L1	4.03
ReLU1	64.76
池化层L2	92.51
卷积层L3	10.19
ReLU2	17.12
池化层L4	21.42
全局平均池化层L5	10.48
总计	220.51

5 结束语

本文提出了基于同态密文转换框架的卷积神经网络推理隐私保护方案。首先本文对于同态卷积神经网络保护框架进行了详细描述，提出了同态友好且高效的同态卷积神经网络模型，设计了卷积层、激活函数、池化层、全局平均池化层的同态运算，并给出同态CNN每层的具体运算和同态操作数。接下来，本文给出了方案的安全性分析。最后，通过实验实现该方案，本文对提出方案在MNIST数据集上实现同态CNN安全推理的时间开销和准确率进行分析评估。实验结果表明，所提方案具有高准确率和低通信开销，性能良好。

参考文献：

[1] GENTRY C. Fully homomorphic encryption using ideal lattices[C]//Proceedings of the forty-first annual ACM symposium on Theory of computing. New York: ACM Press, 2009: 169-178.

[2] BRAKERSKI Z, VAIKUNTANATHAN V. Efficient fully homomorphic encryption from (standard) LWE[C]//Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science. Piscataway: IEEE Press, 2011: 97-106.

[3] BRAKERSKI Z. Fully homomorphic encryption without modulus switching from classical GapSVP[C]//Proceedings of 32nd Annual Cryptology Conference. Berlin: Springer, 2012: 868-886.

[4] FAN J F, VERCAUTEREN F. Somewhat practical fully homomorphic encryption[J]. IACR Cryptol EPrint Arch, 2012, 2012: 144.

[5] GENTRY C, SAHAI A, WATERS B. Homomorphic encryption from learning with errors: conceptually-simpler, asymptotically-faster, attribute-based[C]//Proceedings of 33rd Annual Cryptology Conference. Berlin: Springer, 2013: 75-92.

[6] DUCAS L, MICCIANCIO D. FHEW: bootstrapping homomorphic encryption in less than a second[C]//Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2015: 617-640.

[7] CHILLOTTI I, GAMA N, GEORGIEVA M, et al. TFHE: fast fully homomorphic encryption over the torus[J]. Journal of Cryptology, 2020, 33 (1): 34-91.

[8] CHEON J H, KIM A, KIM M, et al. Homomorphic encryption for arithmetic of approximate numbers[C]//Proceedings of 23rd International Conference on the Theory and Applications of Cryptology and Information Security. Berlin: Springer, 2017: 409-437.

[9] GILAD-BACHRACH R, DOWLIN N, LAINE K, et al. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy[C]//Proceedings of International Conference on Machine learning. New York: ACM Press, 2016: 201-210.

[10] CHABANNE H, WARGNY A D, MILGRAM J, et al. Privacy-preserving classification on deep neural network[J]. IACR Cryptol EPrint Arch, 2017, 2017: 35.

[11] CHOU E, BEAL J, LEVY D, et al. Faster cryptonets: leveraging sparsity for real-world encrypted inference[J]. arXiv Preprint, arXiv: 1811.09953, 2018.

[12] JUVEKAR C, VAIKUNTANATHAN V, CHANDRAKASAN A. Gazelle: a low latency framework for secure neural network inference[J]. arXiv Preprint, arXiv: 1801.05507, 2018.

[13] BOURSE F, MINELLI M, MINIHOLD M, et al. Fast homomorphic evaluation of deep discretized neural networks[C]//Proceedings of 38th Annual International Cryptology Conference. Berlin: Springer, 2018: 483-512.

[14] CHILLOTTI I, JOYE M, PAILLIER P. Programmable bootstrapping enables efficient homomorphic inference of deep neural networks[C]//Proceedings of Cyber Security Cryptography and Machine Learning: 5th International Symposium. Berlin: Springer, 2021: 1-19.

[15] LOU Q, JIANG L. SHE: a fast and accurate deep neural network for encrypted data[J]. arXiv Preprint, arXiv: 1906.00148, 2019.

[16] MISHRA P, LEHMKUHL R, SRINIVASAN A, et al. Delphi: a cryptographic inference service for neural networks[C]//Proceedings of 29th USENIX Security Symposium. Berkeley: USENIX Association, 2020: 2505-2523.

[17] 任艳丽, 余凌赞, 何港, 等. 一种隐私保护的卷积神经网络预测方案[J]. 计算机学报, 2023, 46(8): 1606-1619.
REN Y L, YU L Z, HE G, et al. A scheme of privacy-preserving convolutional neural network prediction[J]. Chinese Journal of Computers, 2023, 46(8): 1606-1619.

[18] LI M H, CHOW S S M, HU S S, et al. Optimizing privacy-preserving outsourced convolutional neural network predictions[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(3): 1592-1604.

[19] LU W J, HUANG Z C, HONG C, et al. PEGASUS: bridging polynomial and non-polynomial evaluations in homomorphic encryption[C]//Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2021: 1057-1073.

[20] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.

[21] YANG J C, YU K, GONG Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 1794-1801.

[22] LOU Q, JIANG L. HEMET: a homomorphic-encryption-friendly privacy-preserving mobile neural network architecture[C]//Proceedings of the International Conference on Machine learning. New York: ACM Press, 2021: 7102-7110.

[23] MOHASSEL P, ZHANG Y P. SecureML: a system for scalable privacy-preserving machine learning[C]//Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2017: 19-38.

[作者简介]

李瑞琪（1993– ），男，黑龙江尚志人，博士，中国民航大学讲师，主要研究方向为全同态加密、格密码学、云计算安全等。

易琴（1998– ），女，四川资阳人，中国民航大学硕士生，主要研究方向为同态加密、隐私保护等。

黄艺璇（1999– ），女，江西新余人，南开大学硕士生，主要研究方向为同态加密、隐私保护等。

贾春福（1967– ），男，河北文安人，博士，南开大学教授、博士生导师，主要研究方向为网络与信息安全、可信计算、恶意代码分析、密码技术应用等。