2024-8-26
计算机科学与探索
Journal of Frontiers of Computer Science and Technology doi: 10.3778/j.issn.1673-9418.2406098

# 同态加密在深度学习中的应用综述

杨洪朝[1,2]，易梦军[1,3]，李培佳[1,3]，张瀚文[1,3]，申富饶[1,3+]，赵健[4]，王刘旺[5]

1. 计算机软件新技术国家重点实验室(南京大学)，南京 210023
2. 南京大学 计算机科学与技术系，南京 210023
3. 南京大学 人工智能学院，南京 210023
4. 南京大学 电子科学与工程学院，南京 210023
5. 国网浙江省电力有限公司电力科学研究院，杭州 310014
+ 通信作者 E-mail: frshen@nju.edu.cn

**摘要：**随着深度学习在各种领域中的广泛应用，数据隐私和安全性问题变得日益重要。同态加密作为一种能够在加密数据上直接进行计算的加密技术，为解决这一问题提供了可能的解决方案。综述了深度学习与同态加密结合的方法，探讨了如何在加密环境中有效应用深度学习模型。首先介绍了同态加密技术的基础知识，涵盖了其基本原理、不同分类（部分同态加密、有限同态加密、全同态机密）以及全同态加密的发展历程。随后详细介绍了深度学习中的关键模型，包括卷积神经网络和 Transformer 模型。在此基础上，探讨了同态加密与深度学习结合的步骤以及如何将深度学习的各个层（如卷积层、注意力层、激活函数层）适配于同态加密环境。然后，重点综述了现有的将卷积神经网络和 Transformer 与同态加密结合的具体方法，探讨了在加密数据上进行深度学习计算的实现方案以及为了提升效率和精度而采用的性能优化策略，并总结了每种方法的优势和局限性。最后，总结了当前研究的进展，并对未来的研究方向进行了展望。

**关键词：**同态加密；深度学习；卷积神经网络；Transformer

文献标志码：A 中图分类号：TP391

# A Survey on the application of homomorphic encryption in deep learning

YANG Hongchao[1,2], YI Mengjun[1,3], LI Peijia[1,3], ZHANG Hanwen[1,3], SHEN Furao[1,3+], ZHAO Jian[4], WANG Liuwang[5]

1. State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing University, Nanjing 210023, China
2. Department of Computer Science and Technology, Nanjing 210023, China
3. School of Artificial Intelligence, Nanjing 210023, China
4. School of Electronic Science and Engineering, Nanjing 210023, China
5. Electric Power Research Institute of State Grid Zhejiang Electric Power Co., Hangzhou 310014, China

**基金项目**：国家电网有限公司总部科技项目资助“数字化安全管控边缘计算装置自主可控关键技术研究及应用”（5700-202319302A-1-1-ZN）。

This work was supported by the science and technology program of State grid Corporation of China(5700-202319302A-1-1-ZN), which is 'Research and Application of Key Technologies for Self control of Digital Security Control edge computing Devices'.

**Abstract:** With the widespread application of deep learning in various fields, data privacy and security issues have become increasingly important. Homomorphic encryption, a technique that allows computations to be performed directly on encrypted data, offers a potential solution to these problems. This paper surveys methods that combine deep learning with homomorphic encryption, exploring how to effectively apply deep learning models in encrypted environments. Firstly, the basics of homomorphic encryption are introduced, covering its basic principles, different classifications (including partially homomorphic encryption, somewhat homomorphic encryption, and fully homomorphic encryption), and the development history of fully homomorphic encryption. Key models in deep learning, such as convolutional neural network and Transformer, are then detailed. Building on this foundation, the steps of combining homomorphic encryption with deep learning and how to adapt various layers of deep learning (e.g., convolutional layers, attention layer and activation function layer) to the homomorphic encryption environments are discussed. Subsequently, existing methods that integrate convolutional neural network and Transformer with homomorphic encryption are focused on, discussing specific implementation schemes for performing deep learning computations on encrypted data and performance optimization strategies employed to enhance efficiency and accuracy. The advantages and limitations of each method are summarized. Finally, current research progress is summarized, and an outlook on future research directions is provided.
**Key words:** Homomorphic Encryption; Deep Learning; CNN; Transformer

随着大数据时代的到来，深度学习技术[1]在图像识别、自然语言处理、语音识别等领域取得了显著的成果，展现了强大的处理和分析能力。然而，深度学习模型的高效训练和推理通常依赖于海量的高质量数据，这些数据往往包含敏感的个人隐私信息。在数据隐私保护和数据共享需求之间，如何实现高效且安全的数据处理成为了亟待解决的问题。同态加密[2]作为一种能够在加密状态下执行计算的技术，为解决这一矛盾提供了新的可能性。

同态加密是一种允许在密文上直接进行算术运算的加密方法，使得在不解密数据的情况下完成复杂的计算成为可能。利用同态加密，数据所有者可以将敏感数据加密后发送给云端或第三方进行处理，处理结果同样是加密形式，只有数据所有者才能解密查看结果。这种特性为数据隐私保护提供了强有力的支持，同时也为深度学习在隐私敏感领域的应用开辟了新路径。

尽管同态加密技术提供了独特的隐私保护优势，但其在深度学习中的应用仍面临诸多挑战。一方面，同态加密在计算复杂度和性能开销方面存在显著劣势，使得直接应用于深度学习模型较为困难。同时，深度学习模型本身的复杂性和计算量大，也对同态加密方案提出了更高的要求。另一方面，由于同态加密仅支持加法和乘法运算，而深度学习模型中通常涉及乘法和加法之外的非线性运算，如激活函数和归一化操作，这使得直接应用同态加密更加复杂。为了克服这些困难，研究人员探索了多种将同态加密与深度学习相结合的方法，以期在保证数据隐私的前提下，仍能实现高效的模型训练和推理。

本文将系统综述同态加密在深度学习中的应用。首先，介绍同态加密的基本概念及其分类方法，并深入探讨全同态加密技术。接着，介绍深度学习中的常见模型，如卷积神经网络（Convolutional Neural Networks，CNN）[3]和 Transformer[4]。随后，分析了同态加密与深度学习结合的步骤以及如何在同态加密环境中调整深度学习模型的各层结构，包括注意力层、激活函数层等。接下来，介绍同态加密在张量数据处理方面的工作并详细探讨同态加密在 CNN 和 Transformer 中的应用方法。最后，总结现有研究成果，展望未来同态加密与深度学习结合的发展方向和潜在应用。

## 1 相关工作

### 1.1 同态加密概述

同态加密是一种基于数学难题的密码学技术，它允许用户直接在加密数据上执行特定的代数运算，而这些运算的结果在解密后与在原始明文数据

上进行相同运算的结果一致，如图 1 所示。这种性质使同态加密成为数据隐私保护的有力工具。同态加密的基本原理可以表示为以下数学表达式:

$$D(E(m1) \otimes E(m2)) = m1 \otimes m2, \#(1)$$

其中$m$为原始数据，$\otimes$为特定的代数运算（加法或乘法），$E$和$D$分别代表加密算法和解密算法。

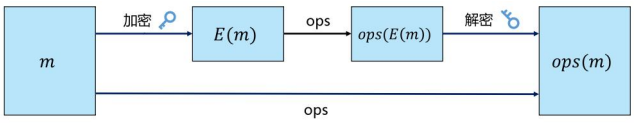



**图 1 同态加密**

Fig.1 Homomorphic encryption

一个完整的同态加密方案通常由四个主要算法组成：密钥生成（KeyGen）、加密（Enc）、评估（Eval）和解密（Dec）。以下是这些算法的基本流程和功能介绍：

(1) 密钥生成（KeyGen）：在密钥生成阶段，系统会根据安全参数生成用于加密和解密的公钥和私钥。假设安全参数为$\lambda$，$pk$和$sk$分别为公钥和私钥，密钥生成过程可以表示为：

$$(pk, sk) = KeyGen(\lambda). \#(2)$$

(2) 加密（Enc）：加密阶段主要使用公钥将明文数据加密为密文数据。输入明文$m$和公钥$pk$，生成密文$c$，加密过程可以表示为：

$$c = Enc(pk, m). \#(3)$$

评估（Eval）：评估阶段在密文上执行指定的计算操作，生成新的密文。输入公钥$pk$，要执行的操作$f$和相关密文$c$，生成密文$c'$，评估过程可以表示为：

$$c' = Eval(pk, f, c). \#(4)$$

(3) 解密（Dec）：解密阶段使用私钥将密文解密为明文，输入密文$c'$和私钥$sk$，使用私钥$sk$对密文$c'$进行解密，恢复计算操作$f$后的结果，解密流程可以表示为：

$$f(m) = Dec(sk, c'). \#(5)$$

### 1.1.1 同态加密分类

同态加密根据其支持的运算类型和次数，可以分为三大类：部分同态加密（Partially Homomorphic Encryption, PHE）、有限同态加密（Somewhat Homomorphic Encryption, SWHE）以及全同态加密（Fully Homomorphic Encryption, FHE）[5]。

部分同态加密（PHE）是一种相对简单的同态加密形式，它仅支持在加密数据上执行某一种运算（加法或乘法）。例如，RSA[6]加密算法和 ElGamal[7]加密算法分别支持乘法和加法运算。PHE 虽然计算效率高，但它仅支持一种特定的计算操作，这限制了其应用范围。

与 PHE 相比，SWHE 能够执行更复杂的计算，支持有限次数的加法和乘法运算。例如，BGN[8]加密方案允许一次乘法和任意次加法运算。SWHE 本身的实际应用可能较为有限，但它在密码学研究中的价值在于为 FHE 的发展提供了一个中间步骤，通过理解和优化 SWHE，研究者们能够更好地探索和完善 FHE 系统。

全同态加密是指可以在不限制次数的情况下，支持任意多次加法和乘法运算的同态加密。这种技术首次由 Craig Gentry 在 2009 年提出，是同态加密领域的一项重大突破。FHE 可以实现对加密数据的任意复杂计算，但其计算开销和实现难度也相对较高。

### 1.1.2 全同态加密技术

全同态加密（FHE）被认为是实现数据隐私保护和计算功能兼顾的理想解决方案。FHE 方案允许在不解密数据的情况下，对加密数据进行任意多次的加法和乘法运算，最终解密的结果与直接对明文进行相同运算的结果一致。

Craig Gentry[9]在 2009 年通过引入 Bootstrapping 技术，提出了第一个实际可行的全同态加密方案。Gentry 的方案基于格理论，首先构建了一个有限同态加密方案，随后利用 Bootstrapping 技术进行“刷新”，即在不解密的情况下去除累积的噪声，从而实现了理论上无限次的计算能力。

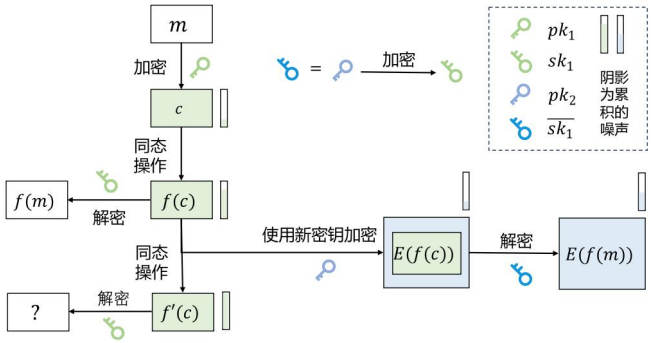



**图 2 Bootstrapping 流程**

Fig.2 Bootstrapping

Bootstrapping 技术的核心思想在于通过一个解密再加密的过程来降低噪声水平。这一过程涉及到使用一个新的密钥对原始密文进行再次加密，以此产生一个噪声更少的新密文。如图 2 所示，假设一个有限同态加密方案有两对公钥和密钥，分别为$(pk_1, sk_1), (pk_2, sk_2)$。明文消息$m$在使用$pk_1$加密后的密文为$c$，密文$c$经过同态操作评估后的密文为$f(c)$，刷新密文$f(c)$噪声的过程如下:

(1) 使用$pk_2$来加密$sk_1$得到

$$\overline{sk_1} = E(pk_2, sk_1). \#(6)$$

(2) 使用$pk_2$来对密文$f(c)$进行加密得到$E(f(c))$，即

$$\bar{c} = E\big(pk_2, f(c)\big). \#(7)$$

(3) 使用加密的密钥$\overline{sk_1}$来对密文$\bar{c}$进行解密，从而得到明文消息$m$在第二个公钥$pk_2$加密下的密文$E(p\widehat{k_2, f}(m))$。该密文与直接使用$pk_2$来对明文消息$f(m)$加密$E(pk_2, f(m))$具有相同的信息，但噪声水平低于密文$f(c)$。

然而，Gentry 的初始方案计算开销巨大，不适用于实际应用。近年来，研究人员在 Gentry 的基础上提出了多种改进的 FHE 方案，研究者们针对不同的技术基础、安全假设和效率特点，提出了多种全同态加密方案。Smart 和 Vercauteren[10]在 2010 年的工作中提出了改进 Gentry 的基于理想格的 FHE 方案。然后，Van Dijk[11]等人在 2010 年引入了基于近似公因数问题的整数域上的 FHE 方案，该方案主要特点是概念上的简单性。随后，Brakerski 和 Vaikuntanathan[12]在 2011 年又提出了另一种基于容错学习问题（Learning With Errors, LWE）问题的 FHE 方案，所提方案具有一定的效率特性。最后，Lopez-Alt[13]等人在 2012 年提出了一种具有高效和标准化特性的 NTRU（Number Theory Research Unit）型 FHE 方案。上述几种加密方案是在不同的数学结构和安全假设下被构造和优化的。因此，后续的类似尝试可以归结为四个主要的 FHE 家族：（1）基于理想格的加密方案；（2）基于整数的加密方案；（3）基于 LWE 的加密方案；（4）基于 NTRU 的加密方案。全同态加密分类图如图 3 所示。

在上述四种类别中，基于 LWE 问题的全同态加密方案逐渐成为研究的热点，因为这些方案不仅具有较强的安全性，而且在效率上有显著提升，如 BFV（Brakerski–Fan–Vercauteren）[14]、CKKS（Cheon–Kim–Kim–Song）[15]和 TFHE（Torus FHE）[16]等。这些方案在计算效率和实用性上取得了显著进展。BFV 方案是一种基于整数的全同态加密方案，适用于精确计算，特别是对整数运算有良好的支持。CKKS 方案是一种支持近似计算的全同态加密方案，适用于浮点数运算和机器学习模型中的小数计算。RNS-CKKS[17]方案是在 CKKS 方案的基础上引入剩余数系统（Residue Number System, RNS）的扩展版本，通过将大整数分解为多个小模数形式进行并行处理，大幅提升了同态运算的效率。RNS-CKKS 不仅保留了 CKKS 对浮点数近似计算的支持，还显著优化了计算性能，使其在处理大规模浮点运算时更加高效，特别适用于对计算速度要求较高的加密场景。TFHE 方案是一种基于环上同态加密的方案，支持高效的位级运算，适用于需要高性能计算的场景。

全同态加密为在保护数据隐私的前提下进行复杂计算提供了强有力的工具。尽管当前 FHE 技术在计算效率和实际应用中仍面临挑战，但随着研究的不断深入和技术的逐步优化，其应用前景十分广阔。

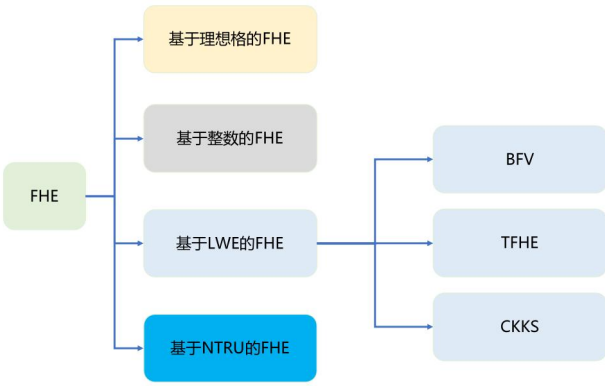



**图 3　全同态加密分类**

Fig.3　Classification of fully homomorphic encryption

## 1.2　深度学习概述

深度学习作为人工智能领域的重要分支，已经在诸多应用场景中展示了其强大的能力。下面将重点介绍卷积神经网络（CNN）和 Transformer 两种常见且重要的深度学习模型。

### 1.2.1　卷积神经网络 CNN

卷积神经网络（CNN）是深度学习中一种专门用于处理具有网格结构数据（如图像）的模型。CNN 通过卷积运算有效提取局部特征，并通过层级结构逐步提取更高层次的特征，广泛应用于图像分类、目标检测、图像分割等任务。

CNN 通常由以下几层组成：

(1) 卷积层（Convolutional Layer）：通过卷积核（或称滤波器）在输入数据上滑动窗口进行局部计算，提取局部特征。

(2) 激活层（Activation Layer）：通常采用非线性激活函数，如 ReLU（Rectified Linear Unit），以引入非线性特性，提高模型表达能力。

(3) 池化层（Pooling Layer）：对卷积层提取的特征进行下采样，减小特征图的尺寸，降低计算量，同时保持重要特征。

(4) 全连接层（Fully Connected Layer）：将提取的特征展平并通过全连接层进行分类或回归任务。

在实际应用中，CNN 的架构往往包含多层的卷积层、池化层和全连接层，以构建深度模型。CNN 的典型架构包括 LeNet[18]、AlexNet[19]、VGG[20]和 ResNet[21]。LeNet 是最早的 CNN 之一，主要用于手写数字识别。AlexNet 在 2012 年 ImageNet 竞赛中取得突破性成果，引入了更深的层级结构和 ReLU 激活函数。VGG 通过使用小卷积核和深层结构在图像分类任务中表现优异。ResNet 引入残差连接，有效解决了深层网络的梯度消失问题，使得网络可以训练得更深。

### 1.2.2　Transformer

Transformer 是一种基于自注意力机制的深度学习模型，最初由 Vaswani 等人在 2017 年提出，用于自然语言处理任务。与传统的 RNN[22]和 CNN 不同，Transformer 能够并行处理序列数据，在处理长距离依赖和大规模数据时具有显著优势。

Transformer 由编码器（Encoder）和解码器（Decoder）组成，如图 4 所示，每个编码器和解码器包含多个注意力模块，每层模块括以下主要组件：

(1) 自注意力机制（Self-Attention Mechanism）：通过计算输入序列中每个元素与其他元素的相关性，动态调整各元素的表示，捕捉长距离依赖关系。常用的自注意力机制包括多头自注意力（Multi-Head Self-Attention, MSA），增强了模型的表示能力。

(2) 前馈神经网络（Feed-Forward Neural Network）：对经过自注意力机制处理的特征进行进一步的非线性变换，通常包括两个全连接层和一个激活函数。

(3) 层归一化（Layer Normalization）：对每一层的输出进行归一化，稳定模型训练，加速收敛。

(4) 残差连接（Residual Connection）：直接将输入添加到输出上，缓解梯度消失问题。

设注意力模块的输入为$z_0$，输出为$y$，注意力层为$MSA$，归一化层为$LN$，前馈网络为$FFN$，则注意力模块首先经过自注意力模块：

$$z_l^{'} = MSA\left(LN(z_{l-1})\right) + z_{l-1}, \#(8)$$

其中$z_{l-1}$是第$l-1$层的输出，$z_l^{'}$是经过自注意力机制处理后的中间表示。然后再通过前馈网络：

$$z_l = FFN\left(LN(z_l^{'})\right) + z_l^{'}, \#(9)$$

$z_l$是经过前馈神经网络处理后的输出表示，最后对输出进行层归一化，得到最终的输出表示：

$$y = LN(z_l). \#(10)$$

Transformer 的典型架构包括 BERT（Bidirectional Encoder Representation from Transformer）[23]、GPT（Generative Pre-trained Transformer）[24]和 ViT（Vision Transformer）[25]。BERT 是基于 Transformer 编码器的双向预训练模型，通过大规模文本数据预训练并在下游任务中微调。GPT 是基于 Transformer 解码器的生成式预训练模型，通过无监督预训练生成高质量文本内容。ViT 将图像划分为若干块，输入 Transformer 编码器处理，展示了其在图像分类和目标检测中的潜力。

CNN 和 Transformer 作为深度学习领域的重要模型，各自在图像处理和自然语言处理等方面展现了强大的能力。在同态加密与深度学习的结合中，如何有效地将这些模型的计算过程与同态加密技术相结合，是实现隐私保护计算的关键。

## 2 深度学习与同态加密

随着深度学习在各个领域的广泛应用，数据隐私保护成为一个重要问题。同态加密技术允许在加密数据上进行计算，为在保护数据隐私的同时利用深度学习提供了一条可行的途径。然而，为了在同态加密环境下正常工作，神经网络的各个层可能需要做出一些调整，以支持加密数据的处理。本章将首先讨论如何将同态加密应用于深度学习以及面临的问题，然后介绍 CNN 和 Transformer 中的网络层，并分析它们在同态加密环境下需要进行的更改。

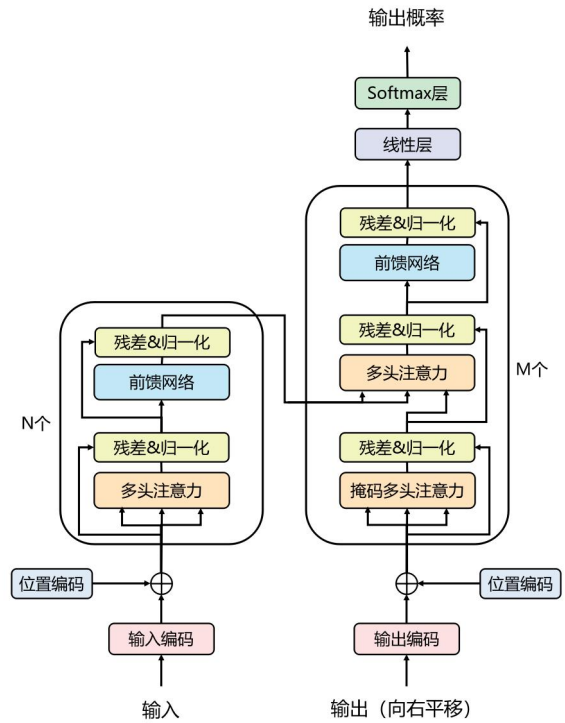



**图 4 Transformer 架构[4]**

Fig.4 The Transformer architecture.

### 2.1 同态加密与深度学习结合的步骤及与挑战

将同态加密应用于深度学习的关键在于同态加密的评估（Eval）阶段。在这一阶段，同态加密在密文上执行指定的计算操作，从而生成新的密文。而在深度学习中应用同态加密，实质上就是在评估阶段对密文执行神经网络的计算，如图 5 所示。

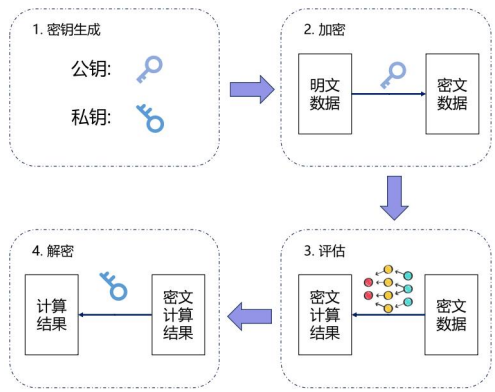



**图 5 同态加密应用于神经网络**

Fig.5 Homomorphic encryption combined with neural networks.

神经网络通常由多个层组成，每一层对输入数据进行一系列变换和计算。这些计算通常包括线性

变换（如矩阵乘法）和非线性变换（如激活函数、归一化、注意力函数）。其中，线性变换在同态加密环境下较为容易实现，因为同态加密本身支持加法和乘法操作。然而，神经网络中的非线性操作是实现网络强大表征能力的关键，但这些非线性操作不能直接应用于同态加密环境，因此需要对这些非线性操作进行改造或近似计算。

除了非线性操作的实现挑战，同态深度学习还面临其他难题。首先是计算复杂度问题，同态加密的计算开销远高于明文计算，复杂的神经网络结构在同态加密环境下的计算更加耗时。数据精度损失也是一个问题，在同态加密下进行的近似计算可能导致数据精度下降，影响应用效果。

## 2.2 神经网络层在同态加密环境下的适应性分析

### 2.2.1 卷积层

卷积层是卷积神经网络（CNN）中的核心组件，其主要功能是通过对输入数据进行滤波操作，从而提取输入数据的局部特征。这一过程依赖于卷积核，它是一个小的二维矩阵，通过在输入数据上滑动并与局部区域进行点乘求和，得到输出特征图中的一个元素。设输入数据为$X$，卷积核为$K$，输出特征为$Y$，卷积操作可以表示为：

$$Y_{i,j} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} X_{i+m,j+n} \cdot K_{m,n}, \#(11)$$

其中$M$和$N$分别是卷积核的高度和宽度，$i$和$j$表示输出特征映射的位置索引。

在同态加密环境下，卷积操作面临的主要挑战是如何在加密数据上有效地进行计算。卷积操作本质上是加权求和，仅用到了乘法和加法，而这两种操作都是同态加密兼容的，因此形式上可以用同态加密支持的加法和乘法进行实现。然而，由于计算效率的限制，需要通过优化策略来提高实际应用中的效率。

### 2.2.2 激活层

激活函数层是深度学习模型中的重要组成部分，用于引入非线性特性，使神经网络能够处理更复杂的函数和模型复杂的决策边界。常见的激活函数包括 ReLU、Sigmoid、Tanh 等。这些函数通过对输入数据进行非线性变换，使得模型能够学习和表示更复杂的模式和特征。ReLU 是目前最常用的激活函数之一，其定义为：

$$ReLU(x) = \max(0, x), \#(12)$$

其主要优点在于计算简单且能够缓解梯度消失问题。

在同态加密环境下，实现激活函数层面临的主要挑战是这些函数通常不是简单的加法和乘法操作，尤其是 Sigmoid 和 Tanh 涉及指数运算，而 ReLU 包含非线性比较运算。因此，需要采用近似方法来实现这些非线性激活函数，比如 Dowlin 等人[26]通过平方函数来进行近似，Lee 等人[27]使用极大极小多项式来近似。

### 2.2.3 池化层

池化层是主要用于下采样（downsampling），以减少特征映射的空间维度，从而降低模型的计算复杂度和防止过拟合。池化操作通过提取局部感受野的统计特征（如最大值或平均值）来保留输入数据的显著信息。常见的池化方法包括最大池化（Max Pooling）和平均池化（Average Pooling）。最大池化通过在局部感受野内选择最大值来进行下采样，平均池化通过计算局部感受野内所有值的平均值来进行下采样。

在同态加密环境下，实现池化层需要特别考虑加密数据的处理方式。由于最大池化涉及比较操作，而同态加密通常不支持直接比较，因此实现上存在一定挑战。相比之下，平均池化只涉及加法和乘法运算，更易于在同态加密环境下实现。

### 2.2.4 全连接层

全连接层的主要功能是将前一层提取到的特征映射到输出空间，通常用于分类、回归等任务。全连接层通过与前一层的每个节点建立连接，并进行线性变换来实现特征的映射。具体来说，全连接层的计算过程包括矩阵乘法和加法操作。设输入特征为$X$，权重矩阵为$W$，偏置向量为$b$，输出为$Y$，则全连接层的计算公式为：

$$Y = WX + b. \#(13)$$

在同态加密环境下，全连接层的实现需要处理加密数据上的矩阵乘法和加法操作。由于同态加密支持加法和乘法操作，全连接层的计算可以在加密域中直接进行，但需要考虑计算效率和加密数据的处理方式。

### 2.2.5 注意力层

注意力机制是近年来在深度学习中广泛应用的一种技术，特别是在自然语言处理（NLP）任务中，如机器翻译、文本生成等。注意力机制通过赋予输入序列中不同部分不同的权重，从而突出重要信息，抑制不重要信息。这种机制可以帮助模型更好地理解和处理长序列数据和复杂结构。注意力机制的核心思想是为输入序列中的每个元素计算一个权重，然后根据这些权重对输入序列进行加权求和。最常用的注意力机制是“点积注意力”（Dot–Product Attention），如图 6 所示，其计算过程可以表示为：

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \#(14)$$

其中$Q,K,V$分别是查询（Query）、键（Key）和值（Value），$d_k$是特征的维度。

在同态加密环境下，实现注意力机制需要处理加密数据上的矩阵乘法、除法和 Softmax 归一化操作。由于同态加密通常不直接支持非线性函数，因此需要采用近似方法[28][29][30]来实现注意力机制的 Softmax 归一化操作。

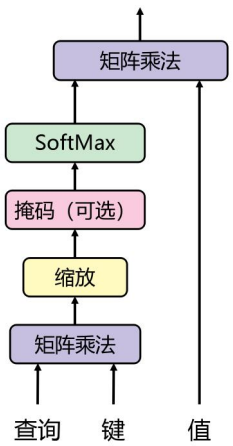



**图 6 Attention 架构[4]**

Fig.6 The Attention architecture

### 2.2.6 归一化层

归一化层的主要作用是通过调整和缩放输入特征，使得模型在训练过程中更加稳定，加速收敛，并且能够提高模型的泛化能力。常见的归一化层包括批量归一化（Batch Normalization）[31]、层归一化（Layer Normalization）[32]和实例归一化（Instance Normalization）[33]等。批量归一化是最常用的归一化技术，主要在每一层的激活值上进行归一化。它通过标准化每个批次的数据，使得其均值为 0，方差为 1，然后引入可学习的参数重新调整标准化后的数据。批归一化可以表示为：

$$y_i = \gamma\frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta, \#(15)$$

其中$\gamma,\beta$是可学习的参数，$x_i,\mu,\sigma$分别是输入数据的特征，均值和方差。

在同态加密环境下，实现归一化层需要特别处理加密数据上的均值、方差的计算和标准化操作。由于同态加密不直接支持除法和开平方运算，因此需要采用近似方法[28][29][30]来实现归一化层的各个步骤。

### 2.2.7 Softmax 层

Softmax 层是深度学习模型中特别常见的一种输出层，尤其在多分类问题中被广泛应用。其主要作用是将输入的未归一化的概率值（logits）转化为归一化的概率分布，从而输出每个类别的概率值。Softmax 函数通过指数函数将输入值进行缩放，再归一化为概率分布，其计算过程如下：

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}, \#(16)$$

其中$p_i$表示输入向量中第$i$个元素对应的概率值。

在同态加密环境下，实现 Softmax 层需要特别处理加密数据上的指数运算和归一化操作[28][29][30]。

### 2.2.8 小结

在同态加密环境下实现深度学习模型中的各种层结构是一个具有挑战性的任务。由于同态加密仅支持加法和乘法操作，需要对传统的深度学习层进行适当的修改和优化，以确保计算在加密域内的可行性和高效性。本章详细介绍了深度学习模型中

的关键层结构，包括卷积层、激活层、池化层、全连接层、注意力层、归一化层和 Softmax 层。在每一部分，我们探讨了这些层的基本工作原理，以及如何在同态加密环境下进行实现和优化，并在表 1 中进行了总结。神经网络层在同态加密环境的主要挑战包括：

(1) 非线性运算的近似：由于同态加密不直接支持非线性运算（如激活函数中的 ReLU、Sigmoid 等），我们需要采用多项式近似等方法来进行计算。

(2) 计算复杂度和效率：同态加密的计算开销较大，需要通过并行计算、数据分块处理等方式来优化计算效率。

(3) 加密数据的归一化：如归一化层和 Softmax 层涉及除法和开平方等运算，需要通过多项式近似或其他协议设计来实现。

**表 1 神经网络层在同态加密环境的适应性分析总结**

Table 1 The adaptability analysis of neural network layer in homomorphic encryption environment

| 神经网络层 | 主要计算操作 | 同态加密环境挑战 | 可能的解决方案 |
|---|---|---|---|
| 卷积层 | 卷积操作（加权求和） | 计算效率受限 | 优化策略，如紧凑打包和并行计算 |
| 激活层 | 非线性变换（ReLU, Simoid, Tanh） | 不直接支持非线性操作，如比较运算和指数运算 | 近似计算：如平方函数、极大极小多项式 |
| 池化层 | 最大池化或平均池化 | 不支持最大池化涉及的比较操作 | 平均池化可直接应用，最大池化需近似或替代方法 |
| 全连接层 | 矩阵乘法与加法操作 | 计算效率和加密数据处理方式 | 需考虑优化策略 |
| 注意力层 | 矩阵乘法、除法、Softmax 归一化 | 不支持直接计算 Softmax 和除法 | 近似计算：多项式逼近方法 |
| 归一化层 | 均值、方差、开平方 | 不支持直接除法和开平方运算 | 近似计算：多项式逼近方法 |
| Softmax 层 | 指数运算与归一化操作 | 不支持直接指数运算和归一化 | 近似计算：多项式逼近方法 |

## 3 同态加密在深度学习中的应用

本章将深入探讨同态加密在深度学习中的应用，首先介绍同态加密在张量数据处理方面的工作，然后重点介绍同态加密与卷积神经网络（CNN）和 Transformer 两种主流的深度学习架构的结合方法。

在深度学习中，张量（Tensor）是表示输入数据、模型参数和中间计算结果的核心数据结构。随着同态加密技术的不断发展，研究人员针对张量数据处理的特性，提出了一系列创新的解决方案，以实现加密状态下高效的张量计算。Feng 等人[34]提出了一种语义安全加密大数据的隐私保护张量分解方法，利用同态加密的特性，使用联邦云为多个用户安全地分解一个加密的张量。此后，Feng 等人[35]设计了一种面向雾计算的隐私保护数据分析模型，该模型结合了张量网络和同态加密的性质，允许用户利用雾计算的存储和分析能力，而不泄露工业环境中的任何敏感数据。Zhang 等人[36][37]提出了 TESLA 和 CANOE，TESLA 旨在在严格的本地差分隐私（LDP）环境下，从包含高噪声的连续输入数据中提取准确的真值，而 CANOE 则用于在地理不可区分性条件下保护用户位置隐私。这些研究展示了同态加密在处理张量数据方面的潜力，并为后续讨论同态加密在深度学习模型中的应用奠定了基础。在接下来的部分中，我们将深入探讨同态加密与卷积神经网络（CNN）和 Transformer 的结合方法。

### 3.1 同态加密在卷积神经网络中的应用

#### 3.1.1 CNN 中同态加密的应用方法

Dowlin 等人[26]提出一种将学习到的神经网络转换为可应用于加密数据的神经网络 CryptoNets 的方法，这使得数据所有者可以将他们的数据以加密的形式发送到云端，而云端能够将神经网络应用到加密数据中进行加密预测，并将预测结果以加密形式返回。CryptoNets 使用了 YASHE[38]加密方案。为了使网络与同态加密兼容，他们对激活函数进行了最低次多项式近似（即平方近似$sqr(z) \coloneqq z^2$）以及

使用平均池化来代替最大池化。该方法在保护数据隐私的同时，能够在 MNIST 数据集上达到 99%的分类精度，每小时可处理 51739 次预测。CryptoNets 的缺点在于其受限于计算复杂度的性能，在更深层次的神经网络模型上表现不佳，准确率下降。

Chabanne 等人[39]的工作优化了 CryptoNets，他们将 Cryptonets 与批归一化（Batch Normalization）层进行了结合，即将多项式逼近与批归一化相结合。他们设计并评估了第一个针对深度大于 2 的神经网络的隐私保护分类方法。他们同样在 MNIST 上验证了他们的方法，并且性能优于 CryptoNets。他们的方法可以应用于具有大量非线性层的神经网络，并且不会降低太多网络的性能。

Hesamifard 等人[40]设计并评估了一种用于深度卷积神经网络的隐私保护分类方法 CryptoDL。他们使用低次多项式来逼近 CNNs 中常用的激活函数（即 ReLU、Sigmoid、Tanh），重点关注了数值分析、标准/修正切比雪夫多项式等近似方法。他们在 MNIST 和 CIFAR−10 数据集上对他们的方法进行了验证，MNIST 上达到了 99.52 %的准确率，CIFAR10 达到 91.5 %的精度。

Sanyal 等人[41]实现了在全同态加密数据上使用复杂模型时的加速，他们主要结合了神经网络的二值化和稀疏化方向的思想，并结合算法工具来加速和并行使用加密数据的计算。

Jiang 等人[42]提出了 E2DM，一种新的矩阵编码方式和对加法、乘法、转置等基本矩阵运算的高效评估策略。E2DM 仅需要线性复杂度来同态运算来计算两个大小为$d \times d$的加密矩阵的乘积，并将基本的矩阵运算扩展到一些高级运算，如转置和矩形矩阵乘法。与 CryptnoNets 方法相比，E2DM 提供了更小的消息大小以及更低的延迟。

Zhou 等人[43]对 CNN 模型的输入数据和权重进行了二值化处理，以实现对加密数据的深度学习的加速。二值化使得 CNN 的同态计算可以以一种高效的方式在二进制域中进行。并且他们还通过设计电路构造了一个高效的池化层，对密文进行比较操作。仿真结果表明，他们所提模型的卷积运算比现有方案至少快 6.3 倍。

Disabato 等人[44]提出了一种创新的分布式深度学习服务（Deep Learning as a Service，DLaaS）设计，在提供基于云的深度学习服务的同时保护用户的敏感数据。所提出的架构使用基于 REST 的客户端服务器框架执行，用于在客户端和服务器之间共享加密数据和发现。他们采用了基于 RLWE 问题的 BFV 方案，在网络层面上面，他们使用平方激活函数代替 ReLU，用平均值代替最大池化。

OBLA 等人[45]指出使用低次多项式作为激活函数会引入误差，从而影响分类任务的准确性。他们介绍了泰勒展开、最佳一致近似、最佳平方近似三种近似方法，并提出了一种加权多项式逼近技术来生成与同态加密兼容更好的激活函数。在三个数据集上的实验结果表明，他们的方法能够提出具有更高或相同精度的与同态加密兼容更好的激活函数。

Lee 等人[27]提出了一种针对 ReLU 和最大池化函数的精确多项式逼近技术，该技术使用小次数的极小极大近似多项式的组合。具体来说，ReLU 的近似多项式为

$$r_\alpha(x) = \frac{x + \alpha p_\alpha(x)}{2}, \#(17)$$

其中$p_\alpha(x)$为 Lee 等人[28]提出的多项式函数。最大池化的近似多项式为

$$N_{\alpha,n}(x_1, x_2, ..., x_n) = \begin{cases} x_1, n = 1; \\ m_\alpha\big(N_{\alpha,k}(x_1, ..., x_k), N_{\alpha,k}(x_{k+1}, ..., x_{2k})\big), \\ \qquad n = 2k; \\ m_\alpha\big(N_{\alpha,k}(x_1, ..., x_k), N_{\alpha,k+1}(x_{k+1}, ..., x_{2k})\big), \\ \qquad n = 2k + 1, \end{cases} \quad (18)$$

其中，$m_\alpha(a, b)$为 CHEON 等人[46]提出的函数。如果将 ResNet、VGG 等深度学习模型中的 ReLU 和最大池化数替换为他们提出的近似多项式，预训练的参数无需重新训练。他们在 ImageNet 上的实验达到了 77.52 %的准确率，这与原始模型 78.31 %的准确率非常接近。

Badawi 等人[47]的工作首次实现了 GPU 加速的同态卷积神经网络（Homomorphic CNN，HCNN）。他们结合了一系列优化技术，如神经网络量化与低精度训练，FHE 方案和参数的优化选择，以及 GPU 加速实现。他们也分别在 MNIST 和 CIFAR-10 上验证了 FCNN，分别取得了 99%和 77.55%的准确率。

Lee 等人[48]指出先前的 PPML 方案并没有使用 Bootstrapping 技术，该技术可以进行连续的同态计算。因此他们在 RNS-CKKS 方案上实现了 ResNet-20 模型，该方案与目前最先进的 ResNet-20 范式几乎相同，只是首次加入了 Bootstrapping 技术。他们在 CIFAR10 数据集上进行了验证，与原始 ResNet-20 的推理结果具有 98.67%的一致性。

Kim 等人[49]提出了更有效的方法来评估 FHE 卷积，无论卷积核大小如何，成本都保持不变，从而在不同的卷积核大小上带来 12-46 倍的时间改进。他们在 CKKS 方案的明文空间中通过算术运算给出了卷积运算的简洁表示，将每个输入打包成环的多项式的系数，并将该方法推广到了批量卷积中。他们还跟 FHE 的 Bootstrapping 程序进行了结合，在 CIFAR10/100 和 ImageNet 数据集上实现了至少 18.9%和 48.1%的 20 层 CNN 分类器的同态评估时间减少。

Zhu 等人[50]提出了第一个基于全同态加密的卷积神经网络（FHE-based Convolution Neural Network，HE-CNN）推理的 FPGA 加速框架。他们基于 FPGA 高层次综合设计流程，设计了参数化的 HE 运算模块，实现了 HE-CNN 层内和层间的资源管理。通过对 HE 运算模块进行复杂的资源和性能建模，所提出的 Fx HENN 框架自动执行设计空间探索，以确定优化的资源供应，并在目标 FPGA 设备上生成给定 HE-CNN 模型的加速器电路。与当前最先进的基于 CPU 的 HE-CNN 推理解决方案相比，FxHENN 实现了高达 13.49 倍的推理加速。

KIM 等人[51]提出了一种深度 HCNN 结构 HyPHEN，该结构包含了最新的卷积算法（RAConv 和 CAConv）、数据打包方法（2D 间隙填充和 PRCR 方案）以及针对 HCNN 结构的优化技术。这样的增强使得 HyPHEN 能够大幅减少内存占用和昂贵的同态操作的数量，例如密文旋转和 Bootstrapping。因此，HyPHEN 将 HCNN CIFAR-10 推理的延迟降到了 1.4 秒（ResNet-20），并在 14.7 秒（ResNet-18）实现了 HCNN ImageNet 推理。

Hu 等人[52]提出了一个新的推理模型框架，该框架利用同态加密的单指令多数据和密文旋转函数以及加法秘密共享（Additive Secret Sharing, ASS）概念来设计和变换 CNN 的每个网络层。该方法在不影响推理精度的前提下，有效地降低了密文膨胀率，同时提高了计算效率，减少了通信量。

### 3.1.2 方法对比与总结

CNN 同态加密方法在数据隐私保护和深度学习模型应用方面展示了重要进展。本文在表 2 中总结了各个研究工作的具体贡献、使用的同态加密方案、存在的缺点以及使用的深度学习数据集。

这些方法采用了不同的同态加密方案，如 YASHE、BGV[53]、TFHE、RNS-CKKS 等，并通过多项式逼近、二值化、分布式架构、GPU 加速、FPGA 加速等技术手段优化 CNN 在加密数据上的性能。尽管在深层网络的处理上仍存在一些局限性和挑战，但在不同数据集上的实验结果表明，这些方法在有效保护隐私的同时，能够实现较高的分类精度和计算效率，为进一步研究和应用奠定了坚实基础。

**表 2 CNN 同态加密方法比较**

Table 2 Comparison of homomorphic encryption methods for CNN

| 方法 | 同态加密方案 | 贡献 | 缺点 | 数据集 |
|---|---|---|---|---|
| **CryptoNets[26] (2016)** | **YASHE** | 第一个可用于加密数据的神经网络 | 非线性层较少时，性能和准确率都比较高，但不适用于深层神经网络 | **MNIST** |
| **Chabanne[34] (2017)** | **BGV** | 将ReLU的多项式逼近与批归一化技术结合，使得深度神经网络可以与同态加密结合 | 分类精度依赖于激活函数的近似性 | **MNIST** |
| **CryptoDL[40] (2017)** | **BGV** | 使用低次多项式来逼近 CNN 中的激活函数 | --- | **MNIST、CIFAR10** |
| **TAPAS[41] (2018)** | **TFHE** | 结合二值化思想来对 CNN 同态加密进行加速 | 仅支持二值化神经网络 | **Cancer、Diabetes、Faces、MNIST** |
| **E2DM[42] (2018)** | **RNS-CKKS** | 提出了新的矩阵编码方式来加速矩阵同态运算 | 局限于简单的矩阵运算 | **MNIST** |
| **Zhou[43] (2020)** | **TFHE** | 在加密数据上进行二值化 CNN 推理，高效池化层 | 全连接层并未二值化，占据了整个模型 51.2%的时间 | **MNIST、Breast Cancer** |
| **Disabato[44] (2020)** | **BFV** | 分布式深度学习同态加密服务架构 | --- | **MNIST、FMNIST** |
| **OBLA[45] (2020)** | **--** | 提出了一种加权多项式逼近技术 | 需要逐层微调权重 | **MNIST、FMNIST、CIFAR-10** |
| **Lee[27] (2023)** | **RNS-CKKS** | 使用极小极大近似多项式的组合来逼近 ReLU 和最大池化的多项式，并在大规模数据集 ImageNet 进行了实验 | 受到高阶多项式（例如，29 )进行近似导致的长延迟限制 | **CIFAR-10、ImageNet** |
| **Badawi[47] (2020)** | **BFV** | 第一个 GPU 加速的同态加密 CNN，结合了神经网络领域的优化技术 | --- | **MNIST、CIFAR-10** |
| **Lee[48] (2022)** | **RNS-CKKS** | 在 RNS-CKKS 方案上实现了 ResNet-20，并首次中加入了 Bootstrapping 技术 | 需要与通道数一样多的 Bootstrapping 操作 | **CIFAR-10** |
| **Kim[49] (2023)** | **RNS CKKS** | 提出了一种卷积核大小无关的全同态加密卷积评估方法 | 限制了网络的灵活性，每个卷积必须与 Bootstrapping 配对 | **CIFAR-10/100、ImageNet** |
| **FxHENN[50] (2023)** | **RNS-CKKS** | 基于全同态加密 CNN 推理的 FPGA 加速框架 | --- | **MNIST、CIFAR-10** |
| **HyPHEN[51] (2023)** | **RNS-CKKS** | 将两种卷积方法与重排序和 2D 间隙填充相结合，显著减少卷积中的同态旋转次数来实现快速推理 | --- | **CIFAR-10、ImageNet** |
| **Hu[52] (2024)** | **CKKS** | 基于同态加密的性质和 ASS 重新设计了 CNN 的网络层，降低了密文膨胀率 | 与明文条件相比，方法的复杂度和计算量仍然较高，仍有改进的空间 | **MNIST** |

## 3.2 同态加密在 Transformer 中的应用

### 3.2.1 Transformer 中同态加密的应用方法

Chen 等人[28]提出了 THE–X，这是一种使预训练 Transformer 模型能够在同态加密环境下进行推理的实用方法。THE–X 通过一个工作流处理 Transformer 网络中的复杂计算，包括所有非多项式函数，如 GELU、Softmax 和 LayerNorm。具体来说，THE–X 使用 ReLU 替换了模型中的 GELU，在用户端执行 ReLU 中的 Max

操作。Softmax 的近似函数为

$$S(x_i) = x_i * T(\sum_j ReLU((\frac{x_j}{2}+1)^3)), \#(19)$$

其中 T 是一个三层的线性神经网络。LayerNorm 的近似方法为

$$y = x \circ \gamma + \beta, \#(20)$$

其中$\gamma, \beta$是可学习的参数，$\circ$是 Hadamard 积。实验结果表明，THE-X 算法在针对不同下游任务进行加密数据推理时，仅表现出微小的性能下降，但在隐私保护方面具有理论上的保障。

Hao 等人[29]提出了一种混合密码框架 Iron，用于 Transformer 隐私推理，不泄露任何关于服务器模型权重或客户端输入的敏感信息。Iron 的主要贡献是为矩阵乘法和复杂的非线性函数（如 Softmax，GELU 激活和 LayerNorm ）提供了几种新的安全协议。具体来说，他们提出了一个定制的基于同态加密的矩阵乘法协议，该协议主要依赖于一种新的紧凑打包技术，实现了$\sqrt{m}$（m 是输出矩阵的行数）的通信加速。其次，设计了 Softmax、GELU 和 LayerNorm 的高效协议，这些协议是建立在 SIRNN 之上的，并且做了一些定制的优化，例如减少 Softmax 中的指数运算开销，简化 GELU 和 LayerNorm。实验结果表明，Iron 与原 SOTA 相比，实现了 3–14 倍通信加速和 3–11 倍的运行加速。

Lu 等人[30]提出了 BumbleBee，一个快速且通信友好的 Transformer 推理框架。他们首先优化了基于同态加密的大矩阵乘法协议，比现有方法减少 80–90%的通信成本。其次，他们提供了一种通用的方法来设计 Transformer 中非线性激活函数的高效和准确的协议，该方法不需要改变 Transformer 模型中的任何组件，因此也不需要任何额外的模型微调。他们在四个 NLP 预训练 Transformer 模型和一个视觉预训练 Transformer 模型和四个数据集上评估了 BumbleBee 的准确性。

Zimerman 等人[54]开创性地提出了第一个多项式 Transformer 模型，并给出了利用 Transformer 对同态加密进行安全推理的证明。他们探索了为同态加密量身定制的 Transformer 结构，以及将运算符转换为其多项式等效的新方法。这一创新使得可以使用语言模型在 WikiText－103 进行安全推理，以及在 CIFAR-100 和 Tiny-ImageNet 进行图像分类。

Liu 等人[55]提出了一个框架来提高基于 Transformer 模型的隐私推理的效率，主要将计算和通信繁重的层或函数转换为密码学友好的层或函数，并在替换过程中微调模型以保留模型精度。该框架在保持近似相同的模型精度的同时，实现了私有推理时间和通信开销的显著减少。

Zheng 等人[56]提出了 Primer，以实现在加密数据上快速而准确的 Transformer。Primer 由一个针对基于注意力的 Transformer 模型优化的混合密码协议构建，同时包括计算合并和 token 优先密文打包等技术。在加密语言模型上的综合实验表明，Primer 取得了当时最好的准确率，并且比之前的方法减少了 90.6 %–97.5 %的推理延迟。

Lee 等人[57]提出了一种高效的基于同态加密的迁移学习算法 HETAL，使用 CKKS 同态加密方案对客户端数据进行加密。HETAL 提出了一种高效的加密矩阵乘法算法，比之前的方法快 1.8 到 323 倍，并且提出了一种覆盖率增加的高精度 Softmax 近似算法。在 5 个基准数据集上的实验结果表明，加密训练造成的准确率损失最大为 0.5%。

Zhang 等人[58]提出了第一个非交互式 Transformer 隐私推理协议 NEXUS。具体来说，他们提出了高效和通信优化的矩阵乘法，采用密文压缩和解压缩策略，并利用矩阵乘法算法中单项式的特殊性质来减少通信开销。他们还提出了一种新颖的 Argmax 算法，将计算复杂度降低到 O ( log m )。

### 3.2.2 方法对比与总结

本文在表 3 中比较了不同 Transformer 同态加密方法，他们采用的同态加密方案包括 CKKS、BFV 等。通过优化矩阵乘法、紧凑打包技术和多项式近似、密文压缩和解压缩等手段，这些方法显著提高了 Transformer 在加密数据上的性能，在处理复杂非线性层（如 GELU、Softmax、LayerNorm）时展示了独特的解决方案，并且在不同数据集上的实验结

果证明了其在保护隐私的同时，实现了较高的准确率和效率。然而，这些方法仍面临一些挑战，例如通信开销和模型微调需求。

**表 3 Transformer 同态加密方法比较**

Table 3 Comparison of homomorphic encryption methods for Transformer

| 方法 | 同态加密方案 | 贡献 | 缺点 | 数据集 |
|---|---|---|---|---|
| **THE-X[28] (2022)** | **CKKS** | 第一个将同态加密与 Transformer 结合的方法 | 每个非线性层的输入会被泄露给客户端，仅在 NLP 任务上进行了验证 | **GLUE、CONLL2003** |
| **Iron[29] (2022)** | **BFV** | 基于紧凑打包技术提出了一种同态加密矩阵乘法协议 | 非线性函数在隐私推理过程中耗时较高 | **GLUE** |
| **BumbleBee[30] (2023)** | **---** | 优化了同态加密大矩阵乘法的效率，显著减少了通信成本 | --- | **GLUE、ImageNet** |
| **Zimerman[54] (2023)** | **CKKS** | 提出了第一个多项式 Transformer | 需要额外的模型训练阶段 | **Wikitext-103、CIFAR-10、Tiny-ImageNet** |
| **Liu[55] (2023)** | **BFV** | 将计算和通信繁重的层或函数转换为密码学友好的层或函数 | 组件替换后需要模型微调 | **GLUE** |
| **Primer[56] (2023)** | **--** | 通过 token 优先打包来代替先验特征优先打包，降低了同态加密的开销 | 局限于 NLP 任务 | **GLUE、SQuAD** |
| **HETAL[57] (2023)** | **CKKS** | 同态加密迁移学习算法，提出了高效的加密矩阵乘法算法和高精度 Softmax 近似方法 | --- | **MNIST、CIFAR10、Face Mask、SNIPS、DermaMNIST** |
| **NEXUS[58] (2024)** | **RNS-CKKS** | 非交互式 Transformer 隐私推理协议，分摊友好矩阵乘法算法和低复杂度 Argmax 算法 | --- | **GLUE** |

## 4 总结与展望

在本文中，我们详细探讨了同态加密在深度学习中的应用，重点关注其与卷积神经网络（CNN）和 Transformer 模型的结合方法。通过对多个前沿研究成果的分析，我们发现同态加密在保护数据隐私的同时，能够有效地支持深度学习模型的推理。对于CNN的同态加密方法，如CryptoNets和CryptoDL，这些方法主要通过低次多项式逼近、矩阵编码和加密计算加速技术来实现。在较浅层的神经网络中，这些方法展示了较高的分类精度和较低的计算延迟。然而，在深层神经网络中，性能和准确率仍需进一步提升。

Transformer 的同态加密方法，如 THE-X、Iron 和 BumbleBee，通过优化矩阵乘法和非线性函数的同态加密计算，实现了 Transformer 模型在加密数据上的推理。这些方法在保持模型准确性的同时，显著减少了计算和通信开销，展示了在自然语言处理（NLP）和图像分类任务中的应用潜力。

未来研究应进一步优化同态加密方案的计算效率，特别是针对深层神经网络和复杂的 Transformer 模型。开发新的加密算法和加速技术（如硬件加速，包括 FPGA、GPU）将是关键方向。同时，提高同态加密方案对不同深度学习模型的兼容性，尤其是处理复杂的非线性层和激活函数的同态加密计算，是实现广泛应用的必要条件。

随着隐私保护需求的增加，同态加密与深度学习的结合方法将在医疗、金融和智能制造等领域展现更多应用潜力。研究应进一步验证这些方法在实际应用场景中的可行性和性能。此外，制定同态加密在深度学习中的应用标准，确保其在实际应用中的安全性和可靠性，也是未来的重要任务。

总之，同态加密在深度学习中的应用仍处于快

速发展阶段，尽管面临诸多挑战，但其在保护数据隐私和实现安全计算方面具有广阔的前景。随着技术的不断进步，同态加密将为深度学习的应用带来更多创新和突破。

## 参考文献：


[1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[2] YI X, PAULET R, BERTINO E, et al. Homomorphic encryption[M]. Springer International Publishing, 2014.

[3] LI Z, LIU F, YANG W, et al. A survey of convolutional neural networks: analysis, applications, and prospects[J]. IEEE transactions on neural networks and learning systems, 2021, 33(12): 6999-7019.

[4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[5] ACAR A, AKSU H, ULUAGAC A S, et al. A survey on homomorphic encryption schemes: Theory and implementation[J]. ACM Computing Surveys (Csur), 2018, 51(4): 1-35.

[6] RIVEST R L, SHAMIR A, ADLEMAN L. A method for obtaining digital signatures and public-key cryptosystems[J]. Communications of the ACM, 1978, 21(2): 120-126.

[7] ELGAMAL T. A public key cryptosystem and a signature scheme based on discrete logarithms[J]. IEEE transactions on information theory, 1985, 31(4): 469-472.

[8] BONEH D, GOH E J, NISSIM K. Evaluating 2-DNF formulas on ciphertexts[C]//Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005, Cambridge, MA, USA, February 10-12, 2005. Proceedings 2. Springer Berlin Heidelberg, 2005: 325-341.

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

[10] SMART N P, VERCAUTEREN F. Fully homomorphic encryption with relatively small key and ciphertext sizes[C]//International Workshop on Public Key Cryptography. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010: 420-443

[11] VAN DIJK M, GENTRY C, HALEVI S, et al. Fully homomorphic encryption over the integers[C]//Advances in Cryptology–EUROCRYPT 2010: 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, French Riviera, May 30–June 3, 2010. Proceedings 29. Springer Berlin Heidelberg, 2010: 24-43.

[12] BRAKERSKI Z, VAIKUNTANATHAN V. Fully homomorphic encryption from ring-LWE and security for key dependent messages[C]//Annual cryptology conference. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 505-524.

[13] López-Alt A, Tromer E, Vaikuntanathan V. On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption[C]//Proceedings of the forty-fourth annual ACM symposium on Theory of computing. 2012: 1219-1234.

[14] FAN J, VERCAUTEREN F. Somewhat practical fully homomorphic encryption[J]. Cryptology ePrint Archive, 2012.

[15] CHEON J H, KIM A, KIM M, et al. Homomorphic encryption for arithmetic of approximate numbers[C]// Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I 23. Springer International Publishing, 2017: 409-437.

[16] 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.

[17] CHEON J H, HAN K, KIM A, et al. A full RNS variant of approximate homomorphic encryption[C]//Selected Areas in Cryptography–SAC 2018: 25th International Conference, Calgary, AB, Canada, August 15–17, 2018, Revised Selected Papers 25. Springer International Publishing, 2019: 347-368.

[18] 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.

[19] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[20] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[22] SALEHINEJAD H, SANKAR S, BARFETT J, et al. Recent advances in recurrent neural networks[J]. arXiv preprint arXiv:1801.01078, 2017.

[23] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[24] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

[25] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[26] GILAD-BACHRACH R, DOWLIN N, LAINE K, et al. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy[C]//International conference on machine learning. PMLR, 2016: 201-210.

[27] LEE J, LEE E, LEE J W, et al. Precise approximation of convolutional neural networks for homomorphically encrypted data[J]. IEEE Access, 2023.

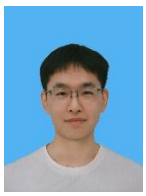
[28] CHEN T, BAO H, HUANG S, et al. The-x: Privacy-

preserving transformer inference with homomorphic encryption[J]. arXiv preprint arXiv:2206.00216, 2022.
[29] HAO M, LI H, CHEN H, et al. Iron: Private inference on transformers[J]. Advances in neural information processing systems, 2022, 35: 15718-15731.
[30] LU W, HUANG Z, GU Z, et al. Bumblebee: Secure two-party inference framework for large transformers[J]. Cryptology ePrint Archive, 2023.
[31] BJORCK N, GOMES C P, SELMAN B, et al. Understanding batch normalization[J]. Advances in neural information processing systems, 2018, 31.
[32] BA J L, KIROS J R, HINTON G E. Layer normalization[J]. arXiv preprint arXiv:1607.06450, 2016.
[33] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance normalization: The missing ingredient for fast stylization[J]. arXiv preprint arXiv:1607.08022, 2016.
[34] FENG J, YANG L T, ZHU Q, et al. Privacy-preserving tensor decomposition over encrypted data in a federated cloud environment[J]. IEEE Transactions on Dependable and Secure Computing, 2018, 17(4): 857-868.
[35] FENG J, YANG L T, ZHANG R, et al. Privacy preserving high-order bi-lanczos in cloud–fog computing for industrial applications[J]. IEEE Transactions on Industrial Informatics, 2020, 18(10): 7009-7018.
[36] ZHANG P, CHENG X, SU S, et al. Task allocation under geo-indistinguishability via group-based noise addition[J]. IEEE Transactions on Big Data, 2022, 9(3): 860-877.
[37] ZHANG P, CHENG X, SU S, et al. Effective truth discovery under local differential privacy by leveraging noise-aware probabilistic estimation and fusion[J]. Knowledge-Based Systems, 2023, 261: 110213.
[38] BOS J W, LAUTER K, LOFTUS J, et al. Improved security for a ring-based fully homomorphic encryption scheme[C]//Cryptography and Coding: 14th IMA International Conference, IMACC 2013, Oxford, UK, December 17-19, 2013. Proceedings 14. Springer Berlin Heidelberg, 2013: 45-64.
[39] CHABANNE H, De WARGNY A, MILGRAM J, et al. Privacy-preserving classification on deep neural network[J]. Cryptology ePrint Archive, 2017.
[40] HESAMIFARD E, TAKABI H, GHASEMI M. Cryptodl: Deep neural networks over encrypted data[J]. arXiv preprint arXiv:1711.05189, 2017.
[41] SANYAL A, KUSNER M, GASCON A, et al. TAPAS: Tricks to accelerate (encrypted) prediction as a service[C]//International conference on machine learning. PMLR, 2018: 4490-4499.
[42] Jiang X, Kim M, Lauter K, et al. Secure outsourced matrix computation and application to neural networks[C]//Proceedings of the 2018 ACM SIGSAC conference on computer and communications security. 2018: 1209-1222.
[43] ZHOU J, LI J, PANAOUSIS E, et al. Deep binarized convolutional neural network inferences over encrypted data[C]//2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom). IEEE, 2020: 160-167.
[44] DISABATO S, FALCETTA A, MONGELLUZZO A, et al. A privacy-preserving distributed architecture for deep-learning-as-a-service[C]//2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020: 1-8.
[45] OBLA S, GONG X, ALOUFI A, et al. Effective activation functions for homomorphic evaluation of deep neural networks[J]. IEEE access, 2020, 8: 153098-153112.
[46] CHEON J H, KIM D, KIM D. Efficient homomorphic comparison methods with optimal complexity[C]// Advances in Cryptology–ASIACRYPT 2020: 26th International Conference on the Theory and Application of Cryptology and Information Security, Daejeon, South Korea, December 7–11, 2020, Proceedings, Part II 26. Springer International Publishing, 2020: 221-256.
[47] Al BADAWI A, JIN C, LIN J, et al. Towards the alexnet moment for homomorphic encryption: Hcnn, the first homomorphic cnn on encrypted data with gpus[J]. IEEE Transactions on Emerging Topics in Computing, 2020, 9(3): 1330-1343.
[48] LEE J W, KANG H C, LEE Y, et al. Privacy-preserving machine learning with fully homomorphic encryption for deep neural network[J]. iEEE Access, 2022, 10: 30039-30054.
[49] KIM D, GUYOT C. Optimized privacy-preserving cnn inference with fully homomorphic encryption[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 2175-2187.
[50] ZHU Y, WANG X, JU L, et al. FxHENN: FPGA-based acceleration framework for homomorphic encrypted CNN inference[C]//2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA). IEEE, 2023: 896-907.
[51] KIM D, PARK J, KIM J, et al. HyPHEN: A Hybrid Packing Method and Its Optimizations for Homomorphic Encryption-based Neural Networks[J]. IEEE Access, 2023.
[52] HU Z, CHEN L, WANG Y, et al. A Secure Convolutional Neural Network Inference Model Based on Homomorphic Encryption[C]//2024 7th World Conference on Computing and Communication Technologies (WCCCT). IEEE, 2024: 17-23.
[53] BRAKERSKI Z, GENTRY C, VAIKUNTANATHAN V. (Leveled) fully homomorphic encryption without bootstrapping[J]. ACM Transactions on Computation Theory (TOCT), 2014, 6(3): 1-36.
[54] ZIMERMAN I, BARUCH M, DRUCKER N, et al. Converting Transformers to Polynomial Form for Secure Inference Over Homomorphic Encryption[J]. arXiv preprint arXiv:2311.08610, 2023.
[55] LIU X, LIU Z. Llms can understand encrypted prompt: Towards privacy-computing friendly transformers[J]. arXiv preprint arXiv:2305.18396, 2023.
[56] ZHENG M, LOU Q, JIANG L. Primer: Fast private transformer inference on encrypted data[C]//2023 60th ACM/IEEE Design Automation Conference (DAC). IEEE, 2023: 1-6.

[57] LEE S, LEE G, KIM J W, et al. HETAL: efficient privacy-preserving transfer learning with homomorphic encryption[C]//International Conference on Machine Learning. PMLR, 2023: 19010-19035.
[58] ZHANG J, LIU J, YANG X, et al. Secure Transformer Inference Made Non-interactive[J]. Cryptology ePrint Archive, 2024.

杨洪朝（2001—），男，河南新乡人，硕士研究生,主要研究领域为同态加密、深度学习。

YANG Hongchao, born in 2001, M.S. candidate. His research interests include homomorphic encryption and deep learning.



易梦军（1997—），男，湖南常德人，博士研究生，主要研究领域为安全计算、联邦学习。

YI Mengjun, born in 1997, Ph.D. candidate. His research interests include secure computation and federated learning.

李培佳（2001—），男，陕西宝鸡人，硕士研究生，主要研究领域为强化学习、机器人控制。

LI Peijia, born in 2001, M.S. candidate. His research interests include reinforcement learning and robotic control.

张瀚文（2001—），男，江苏淮安人，硕士研究生，主要研究领域为联邦学习、同态加密。

ZHANG hanwen, born in 2001, M.S. candidate. His research interest includes federated learning and homomorphic encryption.

申富饶（1973—），男，江苏泰州人，教授，博士生导师，CCF 高级会员，主要研究领域为神经计算、机器人智能。

SHEN Furao, born in 1973, Ph.D., professor, Ph.D. supervisor, senior member of CCF. His research interests include robotic intelligence and neural computing.

赵健（1979—），男，江苏南京人，博士，副教授，IEEE 高级会员，主要研究领域为通信网络、神经计算。

ZHAO Jian, born in 1979, Ph.D, associate professor, senior member of IEEE. His research interests include communication networks and neural computing.

王刘旺（1988—），男，安徽安庆人，博士，高级工程师，主要从事数字化安全管控技术、人工智能技术应用方面的研究工作。

WANG Liuwang, born in 1988, Ph.D, senior engineer. His research interests include applications of digital security management technologies and artificial intelligence technologie.