文章编号:1671-4598(2018)10-0246-04 DOI:10.16526/j.cnki.11—4762/tp.2018.10.053 中图分类号:TP391 文献标识码:A

基于 Alexnet 卷积神经网络的加密芯片模板攻击新方法

郭东昕，陈开颜，张 阳，胡晓阳，魏延海
（陆军工程大学石家庄校区 装备模拟训练中心，石家庄 050003）

摘要：针对经典高斯模板攻击存在的问题，在分析了卷积神经网络方法具有的优势的基础上，提出了一种基于卷积神经网络的加密芯片旁路模板攻击新方法；该方法可以有效地处理高维数据，且可以通过不断地调整网络权值与偏置实现对数据无限逼近，明确分类的精确关系，提高模板刻画精度；最后选取 AT89C52 微控制器（单片机）运行的 AES 加密算法第一轮异或操作为攻击点，与传统的模板攻击进行了对比实验，实验结果表明：虽然在匹配成功率方面稍低于传统的模板攻击，模型结构和超参数仍需要进一步优化，但新方法在处理高维特征点方面较传统的模板攻击具有较大优势。

关键词：加密芯片；卷积神经网络；模板分析；高级加密标准

A New Template Attack Method for Encryption Chip Based on Alexnet Convolutional Neural Network

Guo Dongxin, Chen Kaiyan, Zhang Yang, Hu Xiaoyang, Wei Yanhai
(Shijiazhuang Campus, Army Engineering University, Shijiazhuang 050003, China)

Abstract: Aiming at the problems of classical Gauss template attacks, based on the analysis of the advantages of convolutional neural network, a new method based on convolution neural network for encrypting chip side-channel template attack is proposed. This method can deal with high-dimensional data effectively, and it can achieve infinite approximation of data by adjusting weights and biases constantly, and clarify the precise relationship between classifications, and improve the accuracy of template characterization. Selects the AT89C52 microcontroller (MCU) running AES encryption algorithm first round XOR operation is the point of attack, compared with traditional template attacks. Experimental results show that although the success rate of matching is slightly lower than that of traditional template attacks, the model structure and hyperparameter still need to be further optimized. However, the new method has greater advantages in dealing with high-dimensional feature points than traditional template attacks.

Keywords: encryption chip; convolutional neural network; template analysis; advanced encryption standard

0 引言

在旁路分析[1]领域，可以分为两类，即有模板的旁路分析与无模板的旁路分析，模板攻击[2]是 Chari 等人于 2002 年提出的一种新型的旁路攻击方法，由于该方法具有较好的攻击效果，受到了学术界的普遍关注。传统的旁路攻击方法的效果会受到旁路信号质量的影响，旁路信号噪声较大时，在很多情况下会限制传统的旁路分析方法（如简单能量分析[3]、相关性能量分析[4]、差分能量分析[5-6]等）。然而模板攻击的效果不仅不会受到噪声的影响，反而会得到充分地利用[7]。模板攻击相对于传统的旁路攻击可以从设备泄露的旁路信息中获取利用更多的信息，因此被认为是最强的旁路攻击方式[8]。

由于加密设备在加密过程中泄露的能量与它所处理的数据有关。模板攻击使用多元高斯分布来描述刻画泄露能量迹，然而在处理高维特征数据时，可能会遇到数值计算问题（如奇异矩阵等），并不能很好地处理高维数据。另外，模板攻击使用多元高斯分布作为能量迹概率密度函数并没有进行严格意义上的证明，模型刻画的精度还有提高的空间。

神经网络[9]可以很好地解决这些个问题。神经网络具有较强的自适应能力，可以通过不断地调整权重、偏置等超参数来对训练数据所表达的函数进行无限逼近，从而精准地完成分类任务。同时，神经网络模型属于非线性模型，对于具有复杂关系的数据可以较为灵活有效地拟合。因此，本文针对基于模板的旁路攻击方法存在无法有效应对高维数据、旁路模板刻画精度不够等问题，提出了一种基于 Alexnet[10]卷积神经网络模型的模板攻击新方法。

1 相关概念

1.1 卷积神经网络

当前卷积神经网络[7]在图像分类数据集上的表现较为

收稿日期:2018-03-29； **修回日期**:2018-04-18。
基金项目:国家自然科学基金资助项目(51377170);国家青年科学基金资助项目(61602505)。
作者简介:郭东昕(1993-),男,山西山阴人,在读硕士研究生,主要从事密码安全与旁路分析方向的研究。
陈开颜(1970-),女,辽宁盖县人,副教授,硕导,主要从事密码学方向的研究。

突出，与传统的全连接神经网络最大的区别是减少了需要训练的超参数，提高了训练的速度并较好地解决了神经网络过拟合问题。此外，卷积神经网络模型拥有较好的数据特征提取能力，下面对卷积神经网络分类的实现进行简要介绍。

首先卷积神经网络通过输入层将分类对象数据集 $Di=\{xi, j\}$（i 表示类别，j 表示数据序列）传入模型 Mx（某个特定的卷积模型），其次卷积层 Ci 对输入模型的数据 xi，j 进行特征点提取，生成特征图，Ci 层是模型 Mx 中最为重要的部分，具体的特征提取计算见公式（1）、（2）：

$$f_{i,j,k_1}^{1}=\max(w_{k_1}^{1}x_{i,j}+b_{k_1},0) \tag{1}$$

$$f_{i,j,k_n}^{n}=\max(w_{k_n}^{n}f_{i,j}^{n-1}+b_{k_n},0) \tag{2}$$

其中：x 表示输入数据，w 表示卷积核的权重矩阵，b 表示卷积核的偏移项，k 表示卷积核的大小，n 表示第 n 个卷积层。在卷积层之后，使用池化层 Si 对特征图进行降维，降维有两种计算方式，见公式（3）、（4）：

$$y_i=\frac{1}{n*m}\sum_{i=0}^{n*m}x_i \tag{3}$$

$$y_i=\max(x_1,x_2,x_3,x_4,\cdots,x_{m*n}) \tag{4}$$

公式（3）表示平均池化，公式（4）表示最大池化。池化层可以有效地减少特征维数，提高训练效率。然后通过全连接层 FCi 将提取的特征数据进行重新拟合，有效减少特征信息的损失，最后通过输出层输出结果，完成分类工作。

卷积神经网络区别于其他网络最大的特点即卷积操作，该操作的产生主要有三个方面的因素，分别是：稀疏交互、参数共享、等变表示。

在传统的神经网络的参数矩阵中，每一个参数都描述输入单元与输出单元的交互，而卷积网络具有稀疏交互的特征，即通过远小于输入数据的卷积核来捕获一些较小的特征，减少参数的数量，提高计算效率。参数共享是指相同的参数在多个函数中使用，在传统的神经网络中每一个参数只使用一次，在卷积网络中，卷积运算确保只需要学习一个参数集合而不是每一个位置都学习一个参数集合，极大地减少了需要训练的参数。卷积网络的参数共享的性质使网络层具有了平移等变的特点。即如果函数 $f(x)$ 与 $g(x)$ 满足 $f(g(x))=g(f(x))$，那么函数 f 对于变换 g 具有等变性。在卷积网络中，g 为任意的卷积核平移函数，那么卷积函数对于 g 具有等变性。在处理时间序列数据时，通过卷积可以得到一个由输入中出现不同特征的时刻所组成的时间轴，即输出的表示不变只是在时间上有所变化。

池化也是卷积网路区别于传统的神经网络的操作，操作的目的是通过将某一相邻位置的总体特征来替代这一区域的特征，从而进一步地减少模型训练参数，提高模型的计算的效率。通常情况下，在卷积神经网络中，卷积层与池化层按照一定的比例出现，具体比例会更具具体的实验对象与分类任务有关。卷积神经网络通过不断地卷积与池化操作在稀疏特征与优化计算性能的情况下，不断增强特征提取能力，提高分类效果。

1.2　超参数选择

超参数选择总共分为两种：手动选择与自动选择。手动选择对模型训练人员的要求较高，需要了解相关参数在网络模型中的具体作用。自动选择是通过某种固定模式对参数进行不断调整，不需要了解具体参数的作用，但需要较高的运行时间与计算成本。

通过选择超参数来调整模型的有效容量从而更好地匹配任务。影响有效容量的因素主要有三个，分别是模型的表示容量、学习算法成功最小化训练模型代价函数的能力以及代价函数和训练过程正则化模型的程度。其中学习率是较为重要的超参数，当学习率调节较为适合时，模型的有效容量最高。当学习率过大时，梯度下降会不经意地增加训练误差。相反当学习率太小时，训练周期可能会特别长，对减小训练误差的作用也较弱。

本文实验过程中主要针对模型训练过程中的学习率进行选择。通过训练结果不断调整学习率大小。

1.3　梯度下降

神经网络用于分类，主要是通过对拟合对象的不断迭代从而建立分类对象的数据特征，通过特征对事物完成分类任务。在训练过程中就需要不断地优化函数，即目标函数。在分类任务中称为损失函数，网络模型通过不断地减小损失函数的损失值，从而完成模型的训练。在此过程中就需要用到梯度下降算法。

梯度是对于一个向量求导数，即包含所有偏导数的向量。在多维情况下，临界点是梯度中所有元素都为 0 的点。在具体计算过程中，梯度的方向为增大损失值，为了减小损失值，需要沿着梯度的相反方向。因此该算法被称为梯度下降算法。

1.4　网络架构设计

这里的架构指的是神经网络的整体结构构成，即网络的层数，每层网络的神经元数量以及卷积层与下采样层的配置等。一般来说，网络架构必须通过实验进行验证，通过比较不同的结构对应的实验结果，从而确定网络结构的配置。

当前，神经网络被设计为层的简单链式结构，在实践过程中主要涉及对网络宽度与深度的选择。本文选用的网络架构在图像识别领域具有较好地实验效果，只需要在适应数据结构方面以及实验类别种数进行调整，整体结构依然采用 Alexnet 网络结构。

2　Alexnet 卷积神经网络结构与调整

2.1　Alexnet 卷积神经网络模型概述

Alexnet 模型[10]是 Krizhevsky A 等人于 2012 年提出的深度卷积神经网络模型，并在当年的 ImageNet 大规模视觉识别挑战赛取得第一名的成绩，开启了深度卷积神经网络研发热潮。Alexnet 模型在处理大规模数据方面具有较为高效地特征学习能力，可以较好地掌握高维数据特征，从而

高效地完成分类任务。Alexnet 模型结构如图 1 所示。

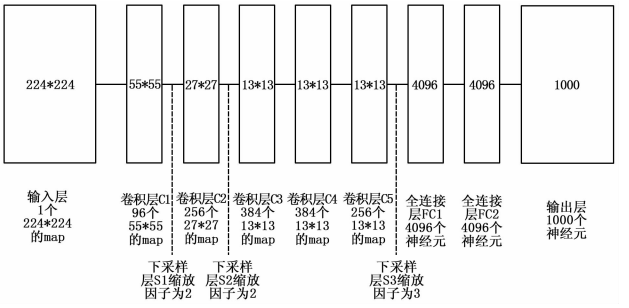

图 1 Alexnet 神经网络模型

以图 1 中输入数据为例，数据的大小为 224×224，即输入层为 224×224，Alexnet 模型共有 5 个卷积层：第一层卷积层 C1 的卷积核大小为 11×11，有 96 个特征图、第二层卷积层 C2 的卷积核大小为 5×5，有 256 个特征图、第三层卷积层 C3 的卷积核大小为 3×3，有 384 个特征图、第四层卷积层 C4 的卷积核大小为 3×3，有 384 个卷积、第五层卷积层 C5 的卷积核大小为 3×3，有 256 个特征图；3 个子采样层：第一个子采样层 S1，缩放因子为 2、第二个子采样层 S2，缩放因子为 2、第三个子采样层 S3，缩放因子为 3；3 个全连接成层：第一个全连接层 F1，有 4096 个神经元、第二个全连接层 F2，有 4096 个神经元、第三个全连接 层，即输出层，有 1000 个神经元。

2.2 Alexnet 卷积神经网络结构优化

将 Alexnet 卷积神经网络模型[10]应用到旁路模板攻击需要对模型结构进行调整，假设 $F=\{T^{(1)}, T^{(2)}, \cdots, T^{(n)}\}$ 是从加密目标设备采集的原始数据集，与传统的模板攻击相同，选取中间值来对数据集 F 进行特征选择。本文选取 AES－128 加密算法第一轮前 8 位密钥异或操作输出的汉明重量作为攻击中间值，中间值的计算方法见公式（5）所示：

$$v_{i,j} = HW(d_i \oplus k_j) \tag{5}$$

式中，d 表示二进制表示的明文值，k 表示加密文明对应的二进制密钥值，v 表示对应异或操作输出值的汉明重量。

这里选取密钥的第一个字节作为攻击目标，根据汉明距离模型，需要构建 9 个不同的模板，也就是不同汉明距离的数值构建不同的模板，将 2562 个数据密钥对映射到 9 个模板。根据前面所述的攻击策略，优化调整后的模型结构如图 2 所示。

根据图 2 所示，以每条数据大小为 1×500 为例进行描述，即输入层为 1×500，Alexnet 模型的 5 个卷积核分别改为 1×11、1×5、1×3、1×3、1×3，特征图数量保持不变。输出层 1000 个神经元调整为 9 个神经元。

2.3 Alexnet 卷积神经网络优化结构的参数

虽然 Alexnet 模型具有较为高效的特征提取能力，但由于应用领域不同，对于数据维数的要求不同，需要对网络结构进行调整，在图像识别领域，数据都为二维形式。但旁路功耗信号为一维数据，因此需要将网络中的卷积核结构以及下采样层进行维数调整。另外由于分类任务中涉及到的类别有 9 种，因此需要将输出层的神经元个数进行修改。

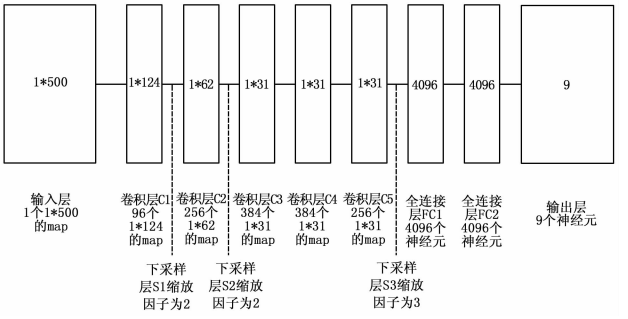

图 2 改进后的 Alexnet 神经网络模型

优化调整后的 Alexnet 卷积神经网络模型需要训练的参数如下：

第一层卷积层 C1 由 96 个特征图构成，卷积核的大小为 11×11，需要训练的参数为（11×11＋1）×96＝11712 个。下采样层 S1，将特征图缩小为原来的 1/2，所以第一层下采样层由 96 个 1×250 的特征图组成。

第二层卷积层 C2 由 256 个特征图构成，卷积核的大小为 5×5，需要训练的参数为为（5×5＋1）×96×256＝638976 个。下采样层 S2，将特征图缩小为原来的 1/2，所以第二层下采样层由 256 个 1×125 的特征图组成。

第三层卷积层 C3 由 384 个特征图构成，卷积核的大小为 3×3，需要训练的参数为（3×3＋1）×256×384＝983040 个。

第四层卷积层 C4 由 384 个特征图构成，卷积核的大小为 3×3，需要训练的参数为（3×3＋1）×384×384＝1474560 个。

第五层卷积层 C5 由 256 个特征图构成，卷积核的大小为 3×3，需要训练的参数为（3×3＋1）×384×256＝983040 个。下采样层 S3，将特征图缩小为原来的 1/3，所以第三层下采样层由 256 个 1×42 的特征图组成。

最后，全连接层将卷积层提取的特征输入全连接式网络，通过输出层得出分类结果，9 个输出神经元分别对应不同的汉明重量。

3 实验设计与结果分析

为了验证新型模板分析的攻击效率，分别进行了传统的基于极大似然函数的模板分析[8]和新型模板分析。实验在 AT89C52 微控制器（单片机）中运行 AES－128 加密算法，选取第一个轮密钥加操作作为攻击点，以前 8 位密钥的汉明重量为例进行分析，由于微控制器总线具有的特点[12]，模板构建采用汉明重量模型，将 8 位操作输出结果分为 9 类，对不同汉明重量的功耗轨迹进行建模。

3.1 传统的模板攻击

传统的模板攻击是基于功耗曲线上的噪声信息服从多元高斯概率分布。在模板攻击中，主要分为三个阶段，分

别是数据特征点提取、特征点模板构建、模板分析。传统的模板攻击不同于神经网络下的模板攻击，需要对每一种类别分别进行模板构建。

在特征点提取和选择过程中，这里通过相关性能量分析方法计算相关操作的相关性系数值，将相关性系数较大的数值所对应的轨迹时刻进行记录，也就是确定了特征点的选取范围。具体的特征点选取采用基于累计均值差的方法[11]，具体公式如下：

$$T_n^i = < r_1, r_2, r_3, \cdots\cdots, r_n > \tag{6}$$

$$m_i = \overline{T^i} \tag{7}$$

$$d_{i,j} = | m_i - m_j | \tag{8}$$

$$D = \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \tag{9}$$

在公式（6）中，T 表示采集样本，n 表示样本点数，i 表示对应类别。公式（7）对同类别的采集样本进行平均化。公式（8）表示对不同类别的采集样本平均值做差。通过公式（9）对公式（8）计算的差异绝对值进行累加求和。

分别选取特征点数量为 5、10、15、20、... 100，使用 10000 条轨迹构建模板，使用 1000 条随机汉明重量轨迹测试集对不同特征点构建的模板进行匹配。模板匹配公式如下：

$$p(T^i;(m,C)_{HW_i}) = \frac{\exp(-\frac{1}{2} \cdot (T^i - m) \cdot C^{-1} \cdot (T^i - m))}{\sqrt{(2 \cdot \pi)^{N_{IP}} \cdot \det(C)}} \tag{10}$$

式中，T^i 表示第 i 条采集样本，m 表示同一类构建模板样本的平均值，矩阵 C 表示构建模板样本的协方差，NIP 表示样本特征点数。通过公式（10）将构建的模板与汉明重量为 0～8 的能量迹进行匹配计算，得到模板匹配概率值。

3.2　新型的模板攻击

基于上面提出的 Alexnet 优化网络模型，同样通过基于累计均值差的方法[11]分别选取特征点数量为 5、10、15、20…100，每个汉明重量使用 1000 条轨迹进行网络模型训练，使用随机汉明重量轨迹测试集对不同特征点构建的模板进行匹配。实验结果如表 1 所示。

由表 1 结果分析可知，在随机明文下，汉明重量为 0 或 8 的能量迹出现的概率最低，为 1/256。传统的模板攻击需要求协方差矩阵的逆，因此，当用于构建模板的特征点数量大于对应建模能量迹数量时，使用传统的模板攻击无法进行模板匹配计算。

与传统的模板攻击相比，基于 Alexnet 深度神经网络的模板攻击对选取特征点的数量没有特定的限制，并且随着选取特征点数量的增加，模型的匹配成功率也相应提高。从表 1 中可以看出，尽管基于 Alexnet 模型的成功匹配略低于传统的模板攻击，但随着特征点数量的增加两种模型的匹配成功率在不断地提高，当特征点数量增加为 50 个时，用于传统模板攻击构建的轨迹为 10000 条，在汉明重量为 0 和汉明重量为 9 两种类别的条数低于 50，因此传统的模板攻击在模板构建构成中协方差矩阵出现计算问题，无法进行模板匹配。而基于 Alexnet 模型的模板攻击不会受到影响。

表 1　特征点数与匹配成功率

选取特征点个数	传统模板攻击匹配成功率	新型模板攻击匹配成功率
5	0.352	0.262
10	0.553	0.339
15	0.644	0.420
20	0.662	0.455
25	0.687	0.494
30	0.705	0.510
35	0.693	0.521
40	0.714	0.561
45	0.727	0.604
50	无解	0.611
55	无解	0.623
60	无解	0.584
65	无解	0.635
70	无解	0.669
75	无解	0.631
80	无解	0.610
85	无解	0.658
90	无解	0.669
95	无解	0.658
100	无解	0.640

目前由于神经网络模型的结构复杂度较低，匹配成功率低于传统的模板攻击，网络模型结构以及参数仍需要进一步调整优化。

4　结论

基于传统的模板攻击选取特征点的数量要小于建模能量迹数量，否则就有可能导致矩阵无法求逆。针对这一问题，本文提出了一种新型的模板攻击方法，虽然目前在在模板匹配成功率方面，由于 Alexnet 深度神经网络模型在训练过程中没有出现过拟合现象，因此模型结构复杂度仍需进一步加强。但在特征点的选取方面要明显优于传统的模板攻击，并且神经网络模型对特征点选取的数量没有具体的限制。

参考文献：

[1] Kocher P C. Timing attacks on implementations of Diffie－Hellman, RSA, DSS, and other systems [A]. N. Koblitz, editor, CRYPTO volume 1109 of LNCS [C]. 1996：104－113.

[2] Chari S, Rao J R, Rohatgi P. Template attacks [A]. CHES2002. LNCS [C]. Springer, Heidelberg. 2003：172－186.

（下转第 254 页）