融合卷积块注意力机制与三元组度量学习的深度侧信道攻击方法

徐 杨[①] 李锴彬*[①] 何星星[②]

[①](西南交通大学信息科学与技术学院 成都 611756)

[②](西南交通大学数学学院 成都 611756)

摘 要：侧信道攻击是密码芯片物理安全的主要威胁之一，利用深度学习技术恢复密钥已成为侧信道攻击领域内研究热点。然而，现有基于深度学习的攻击方法在特征提取阶段往往缺乏对关键泄露区间的聚焦能力。特别是在长波形和高维噪声场景下，模型容易被无关背景噪声干扰，导致特征提取效率低下、猜测熵收敛缓慢。针对上述问题，本文提出一种融合卷积块注意力机制与三元组度量学习的侧信道攻击方法。该方法从泄露特征提取入手，在卷积神经网络中嵌入通道注意力与空间注意力模块，对不同特征通道和时域位置进行自适应加权，以增强泄露相关特征并减弱噪声干扰。在此基础上，引入三元组损失作为优化目标，用于引导注意力加权后的特征在嵌入空间中形成可分的簇结构。实验结果表明，本文方法在公开数据集上均优于度量学习及深度学习方法，其中在ASCAD_f(HW)场景下，攻击性能提升9.4%以上；在ASCAD_r(HW)场景下，本文方法实现猜测熵收敛所需的攻击轨迹数量减少10.7%以上。此外，去同步实验与消融实验进一步证实了该方法具备鲁棒性，并验证了核心组件协同设计的合理性。

关键词：侧信道攻击；深度学习；CBAM；三元组损失；硬件安全

中图分类号：TN918; TP309 **文献标识码**：A **文章编号**：1009-5896(2026)00-0001-11

DOI: 10.11999/JEIT260140 **CSTR**: 32379.14.JEIT260140

1 引言

随着物联网与嵌入式设备的普及，密码芯片的物理安全问题日益凸显。侧信道攻击(Side-Channel Attack, SCA)[1,2]利用设备运行时产生的功耗、电磁辐射和缓存访问时延[3]等泄露信息，能够绕过加密算法的数学安全性直接恢复密钥，已成为硬件安全领域面临的主要威胁之一[4,5]，其应用于嵌入式AI硬件单元的模型提取、对抗样本攻击、模型反转攻击等场景中[6]。其中，模板攻击(Template Attack, TA)[2]通过对目标设备进行预先建模，在信息论意义上被证明是目前最强的SCA手段[7,8]。因此，探索高效、鲁棒的侧信道攻击方法，对评估密码芯片的安全性具有重要的理论意义和实际应用价值。

传统的侧信道攻击方法，如相关功耗分析和经典模板攻击[9]，主要依赖于严格的统计学假设和精确的波形对齐[10]。然而，在实际应用场景中，为了防御侧信道攻击，掩码[11]和随机延迟[12]等对策被广泛部署，这使得采集到的迹线往往伴随着高维噪声、非线性泄露以及时域抖动。在面对此类复杂场景时，传统方法往往因特征工程繁琐、模型假设失效而导致攻击效率急剧下降[13,14]。此外，原始功耗轨迹信噪比低、冗余高维数据掩盖局部泄露以及关键参数设定依赖经验等问题，进一步制约了攻击性能的提升[15]。

近年来，深度学习技术的引入为侧信道攻击带来了突破性的进展[16]。基于深度学习的侧信道攻击正朝着更精细的特征提取与更优的泛化能力方向迅速演进。早期的研究证明了卷积神经网络(Convolutional Neural Network, CNN)能够在无须复杂预处理的情况下提取高阶泄露特征[17]，但面对长波形和高维噪声时，标准卷积操作容易被大量背景噪声分散注意力。为此，研究人员开始将高级特征提取与优化机制引入SCA。Karayalcin等人在近期的研究中指出，注意力机制与特征度量学习已成为突破复杂防御、提升模型判别力的主流趋势[18]。

在特征提取层面，为了在极低信噪比下捕获全局泄露信息，Kulkarni等人打破了传统CNN的限制，将大语言模型中的Transformer架构应用于侧信道攻击，证明了多头自注意力机制的巨大潜力[19]；Huang等人则提出通过融合多维特征与去噪

收稿日期：2026-xx-xx；改回日期：2026-07-01；网络出版：2026-07-13
*通信作者： 李锴彬 likaibin029@icloud.com
基金项目：中央引导地方发展基金(2025ZYDF075)，中央高校基本科研业务费专项资金(2682024ZTPY041, 2682025ZTPY009)，四川省科技计划项目(2024YFHZ0316)，成都市软科学研究项目(2026-RK00-00028-ZF)

Foundation Items: Central Government Guided Local Development Fund (2025ZYDF075), The Fundamental Research Funds for Central Universities (2682024ZTPY041, 2682025ZTPY009), The Science and Technology Planning Project of Sichuan Province (2024YFHZ0316), Chengdu Soft Science Research Project (2026-RK00-00028-ZF)

模型来提升高噪环境下的攻击精度[20]。在特征优化层面，Karayalcin等人基于机制可解释性分析发现，侧信道数据中的密钥相关特征极其稀疏，模型的攻击性能与泛化能力高度依赖于其能否有效聚焦这些关键兴趣点[21]。为了进一步优化特征质量，部分研究者将深度度量学习引入SCA，通过设计特定的损失函数[22]来优化特征在嵌入空间的分布结构，取得了优于传统分类模型的攻击效果。Li等人[23]提出一种基于深度度量学习的轮廓侧信道攻击模型，通过设计归一化提升结构化(Normalized Lifted Structured, NLS)损失、融入标签感知混合距离，有效提升了模型对不同数据集的泛化能力与特征提取效率，为深度度量学习在SCA中的应用提供了新的优化思路。

尽管在注意力机制与度量学习的SCA应用上已取得了显著进展，但现有的前沿研究仍面临几个亟待解决的痛点：首先，现有的多头注意力或Transformer架构往往参数量庞大，在资源受限的嵌入式设备上进行训练开销极大，且在小样本迹线下极易产生过拟合；其次，现有的深度度量学习方法主要聚焦于通过损失函数约束特征距离，却忽视了让网络在特征提取的源头更精准地捕捉泄露信息。侧信道迹线通常包含成千上万采样点，而与密钥相关的有效泄露(Points of Interest, POI)[24]往往只存在于极短的时间窗口内。现有的CNN架构通常采用标准的卷积操作，对所有时间步和通道特征赋予相同的计算权重。这种处理方式导致网络在处理长波形时，容易被大量的背景噪声和无关特征分散注意力，难以在低信噪比环境下精准聚焦微弱的泄露信号，限制了模型的收敛速度和泛化能力。

为了解决上述问题，提出一种融合卷积块注意力机制与三元组的深度侧信道攻击方法。该方法不再单纯依赖损失函数对后端特征的被动约束，而是主动在前端特征提取网络中引入卷积块注意力模块(Convolutional Block Attention Module, CBAM)，赋予网络聚焦能力。核心贡献主要包括：

(1)提出了基于CBAM的自适应特征提取网络。在卷积层之间嵌入通道注意力与空间注意力模块，通过自适应地重新校准特征权重，使网络能够自动识别并关注包含高信噪比泄露的关键时域区间和特征通道，从源头上抑制背景噪声的干扰。

(2)构建了聚焦与优化的协同分析框架。采用三元组损失作为优化目标，利用其强大的特征区分能力，对经过注意力机制筛选后的高质量特征进行优化，显著提升了特征在嵌入空间中的判别性。

(3)验证了方法的高效性与鲁棒性。在公开数据集上的实验表明，该方法不仅在去同步噪声场景下保持了极高的攻击成功率，更重要的是加快了猜测熵的收敛速度，有效降低了评估所需的计算资源和数据量。

2 基础知识

2.1 侧信道攻击与模板攻击原理

SCA的核心在于利用加密芯片在执行算法时所产生的物理泄露与内部中间状态之间的依赖关系来恢复密钥。在建模侧信道攻击场景中，攻击过程通常基于一个普遍的物理泄露假设：芯片的瞬时功耗与当前处理数据的某些物理特征或具体数值存在映射关系。采集到的侧信道迹线建模为如下数学形式：$X=\mathcal{L}(g(D,k))+\varepsilon$。其中，$X\in\mathbb{R}^{1\times L}$表示长度为$L$的侧信道能耗波形向量；$g(\cdot)$表示加密算法中的特定中间值映射函数；$D$为公开的可知数据；$k$为待求解的隐私密钥分量；$\mathcal{L}(\cdot)$描述了中间状态到物理能耗的泄露映射函数；$\varepsilon$则是来自电路环境及采集设备的随机高斯白噪声。

模板攻击是一种基于概率统计的建模攻击方法，通常假设泄露特征服从多元高斯分布。在本文中，利用训练好的深度神经网络作为特征提取器，将攻击迹线l映射为低维嵌入向量$z=f(l)$。

攻击分为两个阶段：模板构建：利用建模数据集计算每个密钥相关中间值k对应的均值向量μ_k和协方差矩阵Σ_k。密钥恢复：对于未知的攻击迹线，计算其属于各个模板的概率。通常采用对数似然函数作为区分器：

$$d(k)=\sum_{j=1}^{N_{attack}}\ln\left(\frac{1}{\sqrt{(2\pi)^D|\Sigma_k|}}\exp\left(-\frac{1}{2}(z_j-\mu_k)^T\cdot\Sigma_k^{-1}(z_j-\mu_k)\right)\right) \tag{1}$$

使得$d(k)$最大的候选密钥即为攻击结果。

2.2 卷积块注意力模块 (CBAM)

CBAM是一种轻量级且通用的注意力机制模块，它可以无缝集成到现有的卷积神经网络架构中，且仅增加极少的参数量和计算开销[25]。CBAM的核心思想是模拟人类视觉系统对重要信息的聚焦过程，通过推断特征图在通道维度和空间维度上的注意力权重，自适应地对特征进行细化。

在侧信道攻击场景中，原始迹线通常包含大量的非相关噪声和冗余时间步。CBAM包含两个独立的子模块：通道注意力模块(Channel Attention Module, CAM)和空间注意力模块(Spatial Attention Module, SAM)。二者通常采用串行连接的方

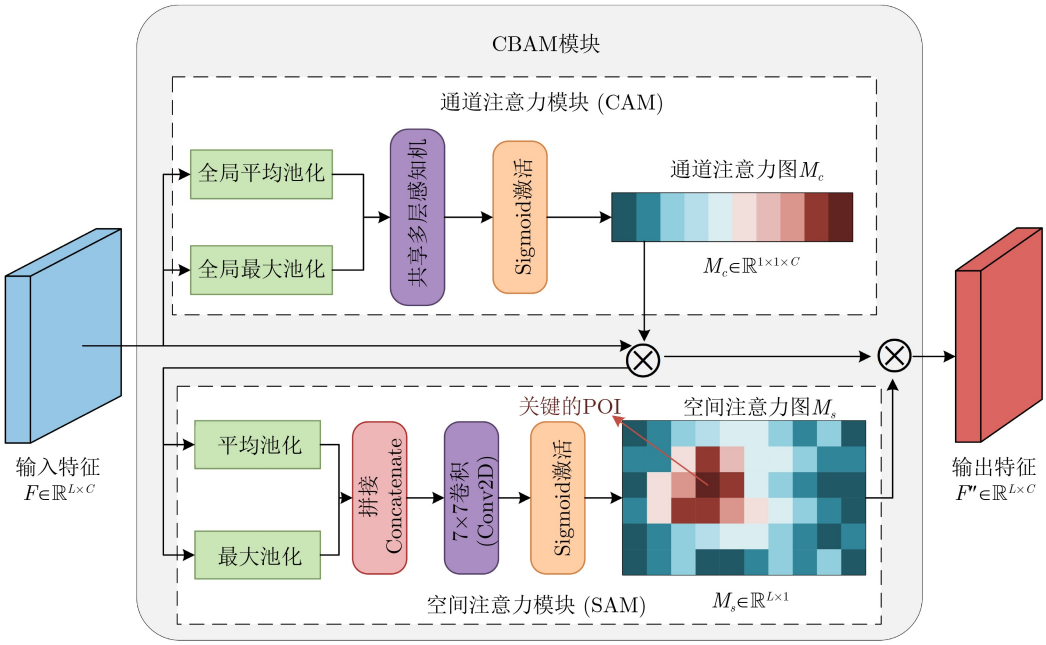

图 1 CBAM模块整体结构示意图

式，依次对输入特征进行关注内容和关注区域的加权处理。

(1)CAM通道注意力机制旨在探索不同特征通道之间的依赖关系。在侧信道信号的卷积特征图中，不同的通道往往响应不同频率或模式的泄露分量。CAM通过显式建模通道间的相关性，赋予包含高信噪比泄露的通道更高的权重，同时抑制包含噪声的通道。为了汇总空间信息，CAM同时采用全局平均池化(Global Average Pooling, AvgPool)和全局最大池化(Global Max Pooling, MaxPool)来压缩特征图的空间维度。假设输入特征为$F \in \mathbb{R}^{L\times C}$，生成的通道描述符分别为$F^c_{avg}$和$F^c_{max}$。这两个描述符随后被送入一个共享的多层感知机(Multilayer Perceptron, MLP)中，以捕获通道间的非线性交互关系。最后，通过Sigmoid激活函数生成通道注意力图$M_c \in \mathbb{R}^{1\times1\times C}$。其计算过程如公式(2)所示：

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{2}$$

其中，σ表示Sigmoid函数，MLP的权重在处理F^c_{avg}和F^c_{max}时是共享的。最终的通道细化特征F'由输入F与广播后的注意力图$M_c(F)$相乘得到。

(2)SAM主要关注特征在时域上的分布，旨在定位关键的POI。在经过通道注意力加权后，SAM进一步在空间维度上筛选信息。与CAM类似，SAM首先沿通道轴方向分别进行平均池化和最大池化，生成两个二维特征图$F^s_{avg} \in \mathbb{R}^{L\times1}$和$F^s_{max} \in \mathbb{R}^{L\times1}$。这两个特征图在通道维度上进行拼接，随后通过一个$7\times1$的卷积层进行融合，最后经由Sigmoid函数生成空间注意力图$M_s \in \mathbb{R}^{L\times1}$。其计算过程如公式(3)所示：

$$M_s(F') = \sigma(f^{7\times1}([AvgPool(F'); MaxPool(F')])) \tag{3}$$

其中，$f^{7\times1}$表示卷积核大小为7的卷积操作。最终的输出特征F''由通道细化特征F'与空间注意力图$M_s(F')$相乘得到。通过这一过程，网络能够有效地抑制非泄露时间段的背景噪声，将计算资源集中在包含敏感信息的时钟周期上。

2.3 三元组损失函数

为了引导注意力网络提取出不仅信噪比高，且在特征空间中具有良好可分性的嵌入向量，本文采用三元组损失作为训练的优化目标。

三元组损失旨在建立特征的类间可分性。通过构建$\{x_a, x_p, x_n\}$三元组，强制锚点样本x_a与同类正样本x_p的欧氏距离比其与异类负样本x_n的距离小一个预设边界α：

$$L_{triplet} = \sum_{i=1}^{N} \left[\|f(x^i_a) - f(x^i_p)\|^2_2 - \|f(x^i_a) - f(x^i_n)\|^2_2 + \alpha\right]_+ \tag{4}$$

其中，$f(\cdot)$表示网络输出的嵌入特征，$||\cdot||_2$表示欧氏距离，$\alpha > 0$为预设的边界参数，用于保证不同类别特征间的分离度。

3 算法流程

本文提出的融合卷积块注意力机制与三元组损失优化的深度侧信道攻击方法，旨在通过前端主动

聚焦、后端紧凑约束的策略，解决复杂噪声环境下泄露特征提取难的问题。算法整体流程如图2所示，主要包含三个核心阶段：数据预处理与注意力特征提取、基于三元组损失的训练、以及基于优化嵌入的模板攻击。

3.1 算法总体架构

为了清晰阐述本文方法的执行逻辑，图2展示了从原始能耗波形输入到最终密钥恢复的端到端技术路线，具体步骤如下：

(1)预处理阶段：输入原始侧信道迹线，进行标准化处理以消除量纲影响，并生成对应的中间值标签。

(2)特征提取与校准阶段：数据进入嵌入了CBAM的一维卷积块。在每一层卷积操作后，利用注意力机制生成权重图，对特征图进行逐元素加权，抑制背景噪声，增强泄露区间的信号强度。

(3)优化阶段：网络输出的低维嵌入向量受三元组损失的约束，拉大不同密钥中间值的类间距离，通过反向传播更新网络权重θ。

(4)攻击阶段：利用训练好的网络提取攻击集迹线的特征，构建多变量高斯模板，计算后验概率并恢复密钥，最后输出最大似然密钥。

3.2 深度特征提取网络结构

特征提取网络的设计直接决定了模型对泄露信息的敏感度。为了在长波形中精准定位POI，本文设计了一种轻量级的注意力卷积神经网络[8]。该网络以一维卷积层(Conv1D)为骨架，在卷积层与池化层之间嵌入CBAM模块，形成卷积-注意力-池化的特征细化结构。网络详细结构参数如表1所示。

芯片能耗波形在时域上极具稀疏性，且特征通

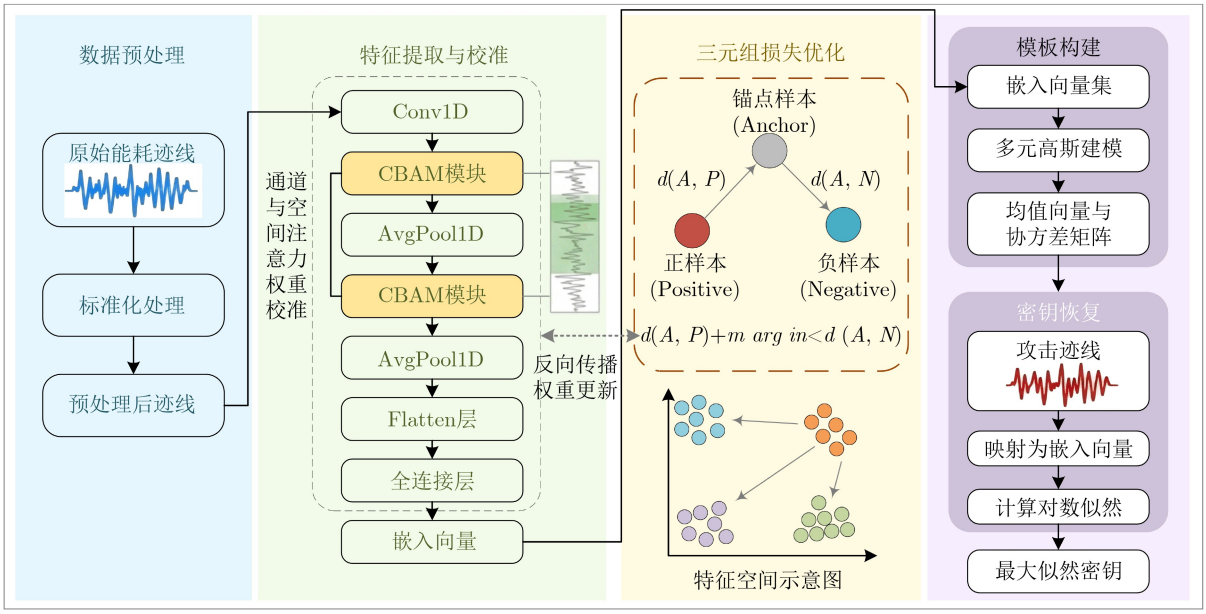

图 2 融合CBAM与三元组损失的侧信道攻击算法流程图

表 1 融合CBAM的特征提取网络结构参数

块	层级	输出尺寸	参数配置	设计动机
Input	Input	$(L,1)$	—	接收原始迹线，L为迹线长度；输入按通道维扩展为一维序列。
Block 1	Conv1D	$(L,64)$	Filters: 64, Kernel: 15, Padding: same, Activation: SELU, Initializer: LeCun Normal	大卷积核捕捉波形整体轮廓；SELU保持数值稳定。
	CBAM	$(L,64)$	Ratio: 8	**关键层：在降采样前对特征进行通道与空间维度的加权校准。**
	AvgPool1D	$(L/15,64)$	Pool size: 15, Stride: 15	平均池化保留能量特征，大幅压缩时间维度。
Block 2	Conv1D	$(L/15,128)$	Filters: 128, Kernel: 3, Padding: same, Stride: 1, Activation: SELU, Initializer: LeCun Normal	小卷积核提取深层局部细节；增加通道数丰富特征语义。
	CBAM	$(L/15,128)$	Ratio: 8	**关键层：进一步聚焦高层语义特征中的有效成分。**
	AvgPool1D	$(L/30,128)$	Pool size: 2, Stride: 2	进一步降维。
Embedding	Flatten	(N_{flat})	—	展平特征图。
	Dense	(16/32)	Activation: Linear	映射至32维嵌入空间，作为度量学习的输入。

道内夹杂大量噪声。CNN由于权重固定，易在层间传递并放大非泄露区间的环境背景噪声，从而掩盖有效泄露信号。而传统的单维度注意力机制(如SE-Net)难以兼顾时空冗余，极易引发噪声干扰。因此，网络引入CBAM模块，通过串行级联CAM与SAM，自适应解析泄露特征的时域区间和特征通道，使模型能够更有针对性地关注泄露相关区域。

此外，当前多头自注意力[5]或Transformer等重型架构虽具备长程建模能力，但其参数量庞杂且计算复杂度随波形长度呈平方级增长。在建模样本有限时，重型网络极易盲目拟合环境高维噪声，引发严重的过拟合。因此，为了避免模型在有限侧信道数据上产生过拟合，同时严格控制基础参数量，网络仅保留两个基础的一维卷积块。其次，模型虽然引入了通道与空间的双重注意力机制，但所采用的CBAM本身是一种极其轻量化的通用模块；其中CAM利用MLP并引入了缩减率以降低隐藏层节点数，而SAM则仅引入单层一维卷积，将计算复杂度严格控制在波形长度的线性阶。这种网络架构使网络在动态聚焦稀疏POI能力的同时，能够有效规避高噪场景下的过拟合风险。

在网络设计中有三大创新点：

(1)CBAM的插入位置：将CBAM模块置于卷积层之后、池化层之前。这是因为池化操作会导致部分空间信息的不可逆丢失。在池化前引入注意力机制，可以确保模型先识别出重要的时间步和通道，赋予其更大的权重，从而保证重要信息在池化后仍被保留，而被抑制的噪声信息则在池化中被进一步过滤。

(2)SELU(Scaled Exponential Linear Unit, SELU)激活函数：为了配合深度度量学习对特征分布的高要求，网络全层采用SELU激活函数[10,22]。配合LeCun Normal初始化[26]，SELU能够实现网络内部的自归一化，有效防止了深层网络训练中梯度的消失或爆炸，保证了特征映射的稳定性。

(3)嵌入维度设定：最终全连接层输出维度设定为32。相较于极低维度(如2维)，32维空间提供了足够的容量来编码复杂的泄露模式；相较于高维度(如128维)，它又避免了稀疏性带来的计算冗余，是性能与效率的平衡点。

3.3 三元组损失函数优化策略

为了充分发挥CBAM提取特征的潜力，采用三元组损失作为优化目标，构建具有良好可分性的特征空间。通过拉大不同密钥中间值的类间距离，确保攻击模型能够区分不同的中间值。设定边界参数$\alpha = 0.1$，使模型学习出具有显著差异的特征表达。

经典的三元组损失函数存在边界参数的敏感度以及对特征空间中噪声和离群点的鲁棒性不足等问题。若直接将其应用于原始功耗波形或浅层CNN提取的粗糙特征，这些特征中包含的大量环境噪声和无关信息会严重干扰距离度量，导致优化过程不稳定甚至陷入局部最优。

为了有效规避上述问题，本文架构在引入三元组损失的同时，通过与轻量化CBAM的结合，规避了上述纯损失函数驱动方法的弊端。在计算三元组损失之前，CBAM模块已经自适应地在通道与空间双重维度完成了对原始能耗波形的动态去噪与信号增强。这意味着，输入到三元组损失计算空间的表征已经具备高信噪比的特征。此时，三元组损失无需再被动地承担低效的特征筛选压力，而是能够最大化发挥其在嵌入空间中的几何拓扑优化能力，确保度量空间的高效与稳定收敛，从而为复杂噪声环境下实现正确密钥的快速恢复提供坚实的结构支撑。

3.4 基于嵌入特征的模板攻击

在攻击阶段，利用训练好的网络$f_\theta(\cdot)$将攻击集迹线映射为低维嵌入向量$Z = f_\theta(X)$。本文采用多变量高斯模型实施攻击：

(1)模板构建：利用建模集，计算每个密钥猜测值对应的中间值类别的均值向量μ_y和协方差矩阵Σ_y。

(2)概率计算与密钥恢复：对于未知的攻击迹线l_{attack}，计算其嵌入向量z属于各个类别的对数似然概率。通过累加所有攻击迹线的对数似然值，选择似然值最大的候选密钥作为最终结果：

$$\hat{k} = \arg\max_k \sum_{i=1}^{N_{attack}} \ln\left(\frac{1}{\sqrt{(2\pi)^D|\Sigma_{y(k)}|}} \cdot \exp\left(-\frac{1}{2}(z_i-\mu_{y(k)})^T\Sigma_{y(k)}^{-1}(z_i-\mu_{y(k)})\right)\right) \tag{5}$$

4 实验结果

为了全面评估本文提出的融合卷积块注意力机制与三元组损失优化的侧信道攻击方法的有效性，本文在两个公开的标准数据集ASCAD和AES_HD上进行了实验。实验采用猜测熵(Guessing Entropy, GE)作为核心评价指标，并统计了猜测熵收敛至1所需的最少迹线数量T_{GE0}以量化攻击效率。所有实验结果均为20次独立重复实验的中位数。

4.1 数据集与实验设置

为了验证算法的有效性，本文选用了侧信道攻击领域常用的ASCAD[27]与AES_HD[12]两个基准数

据集。ASCAD源于运行带掩码防护AES-128算法的8位AVR微控制器。数据集包含两个子集，其中ASCAD_f所有测量使用固定密钥，样本划分为50,000条建模迹线与10,000条攻击迹线；而ASCAD_r中约66%的测量使用随机可变密钥，其余33%使用固定密钥，其中200,000条轨迹用于建模，10,000条固定密钥轨迹用于攻击测试。本研究将攻击点设定为第一轮S盒输出，泄露模型分别采用汉明重量(Hamming Weight, HW)与身份(Identity, ID)两种模型进行对比。AES_HD数据集采集自无防护的FPGA平台，其中使用位于电源线上去耦电容器上方的近场EM探头从FPGA板收集EM迹线。此数据集遵循汉明距离(Hamming Distance, HD)泄漏模型，该模型在AES的最后一轮期间捕获寄存器转换。

为了客观评估不同算法在特征提取方面的本质差异，排除了时间抖动的干扰，基础对比与消融实验均在严格对齐(Desync=0)的数据集上完成。同时，为了验证本文方法在面临迹线不对齐情况下的鲁棒性，进行了不同级别去同步噪声的扩展实验。模型基于TensorFlow2.10框架搭建，并利用NVIDIA GeForce RTX 3060Ti GPU进行加速计算。

4.2 评价指标

Benadjila等人[27]提出多个有关侧信道攻击的评估指标，其中，猜测熵(Guessing Entropy, GE)是目前最广泛认可的性能指标，因此本文采用GE作为衡量攻击成功率的核心指标。GE定义为在多次重复实验中，正确密钥在攻击者预测的密钥排序列表中的平均排名[1]：

$$GE = \frac{1}{N_{exp}} \sum_{i=1}^{N_{exp}} rank(k^*) \tag{6}$$

当GE收敛至1(或0，取决于定义)时，表示攻击者已成功锁定正确密钥。为了更直观地量化攻击效率，本文使用T_{GE0}指标，即猜测熵稳定收敛至1所需的最少攻击迹线数量。T_{GE0}值越小，表明模型提取泄露信息的能力越强，攻击效率越高。

由于深度学习模型的训练过程受到权重初始化等随机因素的影响，单次实验结果可能存在波动。为了消除算法随机性带来的偏差，保证实验结果的统计鲁棒性，报告的所有实验数据均为20次独立重复实验结果的中位数。具体而言，针对每一次超参数设置，我们独立训练并攻击20次，计算每次的T_{GE0}，最终取这20个值的统计中位数作为衡量该方法性能的最终指标[22,24]。

4.3 攻击性能分析

为了评估本文方法(Ours)在不同泄露场景与噪声环境下的特征提取能力，本节在包含固定密钥与随机密钥的ASCAD数据集以及AES_HD上进行了统一评估。我们将本文方法与现有的主流深度学习攻击方法进行了对比，其中经典CNN模型[8]、RL-SCA[10]、FS-SCA[11]、传统度量学习方法(Metric Learning)[22]、NLS[23]、自动化架构搜索方法(AutoSCA-MLP与AutoSCA-CNN)[29]、EL-SCA[30]、CCE[31]以及结合深度特征损失函数的方法(FLR+SoftNN与FLR+Center)[32]均在与本文方法完全相同的实验环境以及统一的评估标准下进行实验。实验结果如表2所示。

如表2所示，在两个公开数据集下，本文方法取得了优异的攻击效果。在ASCAD数据集中，本文方法在ASCAD_f(HW)与ASCAD_f(ID)场景下分别仅需144条和61条攻击迹线即可实现猜测熵中位数的收敛，相较于经典CNN减少了约51.0%/68.0%，相较于Metric Learning方法减少了约9.4%/5%。在ASCAD_r(HW)中，本方法需176条迹线，相较于RL-SCA方法减少了约80.7%，相较于Metric Learning方法减少了约10.7%。在ASCAD_r(ID)数据集上，本文方法需137条迹线，相较于Metric Learning、AutoSCA-CNN以及CCE等方法展现了更优的攻击性能。与此同时，在AES_HD数据集中，RL-SCA方法所需迹线数为4415条，基于多头注意力的CNN模型所需迹线数为2068条。

表 2 攻击性能对比

方法	ASCAD_f		ASCAD_r		AES_HD
	HW	ID	HW	ID	HW
MHA[5]	—	—	—	—	2174
DCMHA[5]	—	—	—	—	2068
CNN[8]	294	191	—	—	—
RL-SCA[10]	906	242	911	490	4415
FS-SCA[11]	532	—	538	78	—
LMA-SCA[19]	—	87	—	78	—
Metric Learning[22]	159	64	197	188	1768
NLS[23]	—	—	—	—	1252
SACNN[28]	298	—	212	—	—
AutoSCA-MLP[29]	447	120	617	3481	—
AutoSCA-CNN[29]	539	257	496	2975	—
EL-SCA[30]	—	—	470	105	—
CCE[31]	>2000	963	2840	>3000	—
FLR+SoftNN[32]	832	716	2592	>10000	—
FLR+Center[32]	790	728	3867	9681	—
CNN-fusion[33]	—	149	—	—	1116
Ours	**144**	**61**	**176**	**137**	**1219**

在引入深度度量学习的方法中，传统的Metric Learning[22]和NLS[23]方法所需迹线数分别为1 768条和1 252条。而本文方法所需迹线数为1 219条。为了更直观地展示本方法在攻击过程中的动态收敛特性，图3和图4绘制了本方法在两个数据集及不同泄露模型下的猜测熵随攻击迹线数量变化的曲线。

结合图3与图4的猜测熵中位数收敛曲线可以进一步发现，在不同数据集与泄露模型下，本文方法的收敛过程均表现出下降趋势。这一全局表现直接验证了本文协同架构设计的有效性：CBAM模块的空间与通道注意力机制能够自适应地定位长波形中的有效泄露时刻，从源头上抑制了大量无关时间步与高维背景噪声的干扰；而三元组损失则为经过提纯的高质量特征提供了紧致化的分布约束。两者的优势互补，确保了模型在各类复杂的物理特征与加密配置下，均能实现正确密钥的高效、稳健恢复。

4.4 去同步迹线结果

为验证模型在时钟抖动和波形未对齐等真实物理环境下的鲁棒性，本节在测试集中引入最大偏移量为50和100的随机时间漂移(Desync=50, 100)进行验证，这两种偏移量是评估鲁棒性的标准基准设置，分别代表了中度与重度的波形不对齐现象，足以覆盖大多数真实物理设备中由时钟抖动或随机延迟防御引发的去同步范围。同时，本节将所对比的传统度量学习(Metric Learning)方法在与本文方法相同的实验环境和一致的去同步配置下进行实验。

实验结果如表3所示，随着去同步程度加剧，恢复密钥所需的迹线数量整体呈上升趋势。在Desync=50时，本文方法在ASCAD数据集的ASCAD_f(HW)场景下需240条迹线即可实现猜测熵稳定收敛，优于Metric Learning方法；在ASCAD_r(HW)和ASCAD_r(ID)场景以及AES_HD数据集上，本文方法也表现出较低的迹线需求。在Desync=100的去同步条件下，传统度量学习方法与本文方法在AES_HD数据集上均未能在最大测试迹线数范围内引导猜测熵中位数收敛至1，但本文方法在ASCAD_f(HW)和ASCAD_r(HW)场景下仍低于传统度量学习方法所需迹线数量。

4.5 消融实验

为了深入探究本文所提模块的具体贡献与协同机制，本节以标准一维CNN为基线模型(Baseline)在两个数据集上开展消融实验，结果如表4所示。首先，Baseline在各数据集上均需要大量的攻击迹线才能实现猜测熵收敛，在ASCAD_r(ID)场景下甚至超出10 000条仍未收敛。当在基线模型中单独引入CBAM模块(Baseline+CBAM)或单独引入三元组损失(Baseline+Triplet)时，所需迹线数量均显著下降，这分别证明了卷积块注意力机制在时空去噪和度量学习在特征聚类中的有效性。本文的方

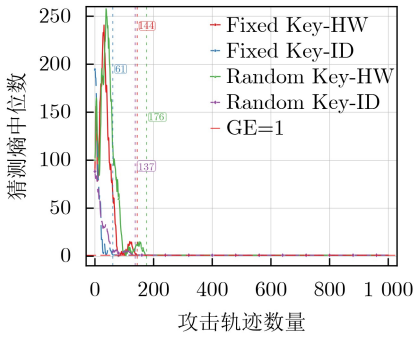

图 3 ASCAD数据集猜测熵收敛曲线

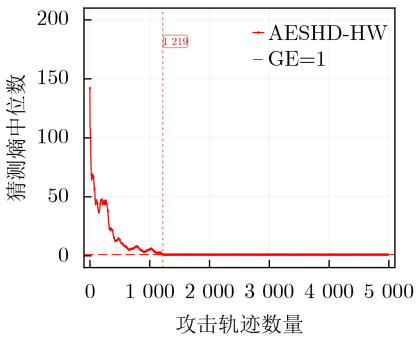

图 4 AES_HD 数据集猜测熵收敛曲线

表 3　去同步迹线结果

Desync	方法	ASCAD_f		ASCAD_r		AES_HD
		HW	ID	HW	ID	HW
50	Metric Learning[22]	251	191	2 251	3 385	4 662
	Ours	240	286	1 898	3 355	4 521
100	Metric Learning[22]	382	582	6 386	9 932	—
	Ours	357	610	6 227	9 973	—

表 4　消融实验结果

方法	ASCAD_f		ASCAD_r		AES_HD
	HW	ID	HW	ID	HW
Baseline	2 064	506	5 841	>10 000	3 051
Baseline+CBAM	1 534	448	4 926	>10 000	2 656
Baseline+Triplet	153	67	197	276	4 059
Ours_post-pool	157	92	284	360	2 805
Ours_Relu	281	904	825	1 171	4 844
Ours_dim2	>10 000	>10 000	>10 000	>10 000	5 994
Ours_dim128	2 631	357	9 287	4 140	>10 000
Ours	**144**	**61**	**176**	**137**	**1 219**

法(Ours)通过结合两者，在所有测试场景下均达到了最优性能。这一结果验证了本文方法的协同架构能够相互促进，从而在复杂噪声环境下最大化特征提取效率。

此外，实验进一步验证了网络结构中关键组件设定的合理性。将CBAM移至池化层之后(Ours_post-pool)会导致性能明显衰退，验证了池化操作会导致部分空间信息的不可逆丢失，而在池化前进行注意力加权，能够确保包含关键泄露的重要时间步在降采样过程中被优先保留；将全层SELU替换为ReLU(Ours_Relu)后，模型在AES_HD上的迹线需求增长至4844条，证明了SELU的自归一化特性是保障高噪环境下数值稳定、防止模型崩溃的关键。

针对网络输出的嵌入向量维度，实验对比了2维、32维以及128维的性能差异。结果显示，过低的维度(Ours_dim2)会引发严重的特征容量受限与信息丢失，导致模型在所有ASCAD场景下均完全失效。反之，过高的维度(Ours_dim128)则容易引入多余的稀疏特征与计算冗余，极大地增加了模型在有限建模样本下盲目拟合环境背景噪声的过拟合风险，导致在AES_HD上无法收敛。相比之下，最终设定的32维嵌入空间在信息容量与计算效率之间取得了最佳平衡，确保了模型在各类数据集上的泛化能力与高效性。

5 结论

本文提出了一种融合CBAM与三元组损失优化的深度侧信道攻击方法。该方法通过在CNN中嵌入通道与空间注意力模块，赋予了网络自适应校准特征权重的能力，从源头抑制了背景噪声；同时利用三元组损失约束嵌入空间，强制异类特征分离，显著增强了特征的判别性。在ASCAD和AES_HD数据集上的实验表明，该方法在猜测熵收敛速度和攻击成功率上均优于现有CNN及传统度量学习模型。去同步实验与消融实验进一步证实了该方法优异的鲁棒性，并验证了网络核心设计的合理性与协同效应。

未来的研究工作将从两个方面进一步推进。首先，将探索更加轻量化的注意力机制，以降低模型在资源受限嵌入式设备上的部署成本；其次，将重点提升模型对更剧烈去同步扰动的适应能力。在随机时间漂移进一步增大时，本文方法可能出现猜测熵难以稳定收敛的现象。因此，后续研究将重点关注如何提升模型在更强去同步扰动下的稳定性和鲁棒性，使其在更复杂的物理设备评估场景中仍能保持有效的攻击性能；同时进一步拓展该方法在小样本场景下的应用，为密码芯片安全性分析提供更全面的技术支撑。

参 考 文 献

[1] MANGARD S, OSWALD E, and POPP T. Power Analysis Attacks: Revealing the Secrets of Smart Cards[M]. New York: Springer, 2007.

[2] CHARI S, RAO J R, and ROHATGI P. Template attacks[C]. 4th International Workshop on Cryptographic Hardware and Embedded Systems, Redwood Shores, USA, 2002: 13–28. doi: 10.1007/3-540-36400-5_3.

[3] 郑帅, 徐向荣, 肖利民, 等. 面向缓存侧信道攻击防护的快速刷写技术[J]. 电子与信息学报, 2025, 47(9): 3178–3186. doi: 10.11999/JEIT250471.

ZHENG Shuai, XU Xiangrong, XIAO Limin, *et al*. Mitigating cache side-channel attacks via fast flushing mechanism[J]. *Journal of Electronics & Information Technology*, 2025, 47(9): 3178–3186. doi: 10.11999/JEIT250471.

[4] LERMAN L, POUSSIER R, BONTEMPI G, *et al*. Template attacks vs. machine learning revisited (and the curse of dimensionality in side-channel analysis)[C]. 6th International Workshop on Constructive Side-Channel Analysis and Secure Design, Berlin, Germany, 2015: 20–33. doi: 10.1007/978-3-319-21476-4_2.

[5] 蒋玲腊, 陈文, 孙伟, 等. 融合动态可组合多头注意力的深度学习侧信道分析方法研究[J/OL]. 计算机科学, 2025: 1–15. https://link.cnki.net/urlid/50.1075.TP.20251129.1824.004, 2025.

JIANG Lingla, CHEN Wen, SUN Wei, *et al*. Research on a deep learning-based side-channel analysis method with dynamically composable multi-head attention[J/OL]. *Computer Science*, 2025: 1–15 https://link.cnki.net/urlid/50.1075.TP.20251129.1824.004, 2025.

[6] 金诚斌, 高宜文, 高锐, 等. 嵌入式AI硬件单元的侧信道分析方法概述[J]. 密码学报(中英文), 2025, 12(4): 729–751. doi: 10.13868/j.cnki.jcr.000791.

JIN Chengbin, GAO Yiwen, GAO Rui, *et al*. Systematic study on physical side-channel attack against embedded AI hardware units[J]. *Journal of Cryptologic Research*, 2025, 12(4): 729–751. doi: 10.13868/j.cnki.jcr.000791.

[7] WU Lichao and PICEK S. Remove some noise: On pre-processing of side-channel measurements with autoencoders[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2020, 2020(4): 389–415. doi: 10.13154/tches.v2020.i4.389-415.

[8] ZAID G, BOSSUET L, HABRARD A, *et al*. Methodology for efficient CNN architectures in profiling attacks[J]. *IACR Transactions on Cryptographic Hardware and Embedded*

Systems, 2020, 2020(1): 1–36. doi: 10.13154/tches.v2020.i1.1-36.

[9] LU Xiangjun, ZHANG Chi, CAO Pei, *et al*. Pay attention to raw traces: A deep learning architecture for end-to-end profiling attacks[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2021, 2021(3): 235–274. doi: 10.46586/tches.v2021.i3.235-274.

[10] RIJSDIJK J, WU Lichao, PERIN G, *et al*. Reinforcement learning for hyperparameter tuning in deep learning-based side-channel analysis[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2021, 2021(4): 677–707. doi: 10.46586/tches.v2021.i3.677-707.

[11] PERIN G, WU Lichao, and PICEK S. Exploring feature selection scenarios for deep learning-based side-channel analysis[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2022, 2022(4): 828–861. doi: 10.46586/tches.v2022.i4.828-861.

[12] HE Pengfei, ZHANG Ying, GAN Han, *et al*. Side-channel attacks based on attention mechanism and multi-scale convolutional neural network[J]. *Computers and Electrical Engineering*, 2024, 119: 109515. doi: 10.1016/j.compeleceng.2024.109515.

[13] 张倩, 高宜文, 刘月君, 等. 基于循环展开结构的抗侧信道攻击SM4 IP核设计[J]. 密码学报(中英文), 2025, 12(3): 645–661. doi: 10.13868/j.cnki.jcr.000786.
ZHANG Qian, GAO Yiwen, LIU Yuejun, *et al*. Design of SM4 IP cores against side-channel attacks based on loop unrolling architecture[J]. *Journal of Cryptologic Research*, 2025, 12(3): 645–661. doi: 10.13868/j.cnki.jcr.000786.

[14] PICEK S, HEUSER A, JOVIC A, *et al*. Side-channel analysis and machine learning: A practical perspective[C]. 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, USA, 2017: 4095–4102. doi: 10.1109/IJCNN.2017.7966373.

[15] 罗玉玲, 徐海洋, 欧阳雪, 等. 高效侧信道分析: 从协同去噪到自适应B样条降维[J]. 电子与信息学报, 2026, 48(3): 1354–1365. doi: 10.11999/JEIT251047.
LUO Yuling, XU Haiyang, OUYANG Xue, *et al*. High-efficiency side-channel analysis: From collaborative denoising to adaptive B-spline dimension reduction[J]. *Journal of Electronics & Information Technology*, 2026, 48(3): 1354–1365. doi: 10.11999/JEIT251047.

[16] LI Kaibin, LIANG Yihuai, ZHOU Zhengchun, *et al*. HypSCA: A hyperbolic embedding method for enhanced side-channel attack[R]. Paper 2025/1199, 2025.

[17] CAGLI E, DUMAS C, and PROUFF E. Convolutional neural networks with data augmentation against jitter-based countermeasures: Profiling attacks without pre-processing[C]. The 19th International Conference on Cryptographic Hardware and Embedded Systems, Taipei, China, 2017: 45–68. doi: 10.1007/978-3-319-66787-4_3.

[18] KARAYALCIN S, KRČEK M, and PICEK S. SoK: Deep learning-based side-channel analysis trends and challenges[J/OL]. *IACR Cryptology ePrint Archive*, 2025(1309). **(查阅网上资料, 未找到本条文献信息, 请确认)**.

[19] KULKARNI P, VERNEUIL V, PICEK S, *et al*. Order vs. chaos: A language model approach for side-channel attacks[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2024, 2024(1): 1–32. **(查阅网上资料, 未找到本条文献信息, 请确认)**.

[20] HUANG Hai, WU Jinming, TANG Xinling, *et al*. Deep learning-based improved side-channel attacks using data denoising and feature fusion[J]. *PLoS One*, 2025, 20(4): e0315340. doi: 10.1371/journal.pone.0315340.

[21] KARAYALCIN S, KRČEK M, and PICEK S. Interpreting emergent features in deep learning-based side-channel analysis[C]. The Thirty-ninth Annual Conference on Neural Information Processing Systems, San Diego, USA, 2025.

[22] WU Lichao, PERIN G, and PICEK S. The best of two worlds: Deep learning-assisted template attack[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2022, 2022(3): 413–437. doi: 10.46586/tches.v2022.i3.413-437.

[23] LI Kaibin, LIANG Yihuai, MENG Hua, *et al*. Deep metric learning-based side-channel analysis with improved robustness and efficiency[J]. *Applied Intelligence*, 2025, 55(10): 712. doi: 10.1007/s10489-025-06586-z.

[24] 谢豪, 田曦, 廖威, 等. 基于射频侧信道的密码芯片安全测评方法[J]. 密码学报(中英文), 2024, 11(3): 706–718. doi: 10.13868/j.cnki.jcr.000703.
XIE Hao, TIAN Xi, LIAO Wei, *et al*. Cryptographic chip security evaluation method based on radio frequency side channel[J]. *Journal of Cryptologic Research*, 2024, 11(3): 706–718. doi: 10.13868/j.cnki.jcr.000703.

[25] WOO S, PARK J, LEE J Y, *et al*. CBAM: Convolutional block attention module[C]. 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19. doi: 10.1007/978-3-030-01234-2_1.

[26] LECUN Y, BOTTOU L, ORR G B, *et al*. Efficient backProp[M]. ORR G B and MÜLLER K R. Neural Networks: Tricks of the Trade. Berlin: Springer, 1998: 9–50. doi: 10.1007/3-540-49430-8_2.

[27] BENADJILA R, PROUFF E, STRULLU R, *et al*. Deep learning for side-channel analysis and introduction to ASCAD database[J]. *Journal of Cryptographic Engineering*, 2020, 10(2): 163–188. doi: 10.1007/s13389-019-00220-8.

[28] HAJRA S, ALAM M, SAHA S, *et al*. On the instability of softmax attention-based deep learning models in side-

channel analysis[J]. *IEEE Transactions on Information Forensics and Security*, 2024, 19: 514–528. doi: 10.1109/TIFS.2023.3326667.

[29] WU Lichao, PERIN G, and PICEK S. I choose you: Automated hyperparameter tuning for deep learning-based side-channel analysis[J]. *IEEE Transactions on Emerging Topics in Computing*, 2024, 12(2): 546–557. doi: 10.1109/TETC.2022.3218372.

[30] PERIN G, CHMIELEWSKI Ł, and PICEK S. Strength in numbers: Improving generalization with ensembles in machine learning-based profiled side-channel analysis[J]. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2020, 2020(4): 337–364. doi: 10.13154/tches.v2020.i4.337-364.

[31] KERKHOF M, WU Lichao, PERIN G, *et al*. No (good) loss no gain: Systematic evaluation of loss functions in deep learning-based side-channel analysis[J]. *Journal of Cryptographic Engineering*, 2023, 13(3): 311–324. doi: 10.1007/s13389-023-00320-6.

[32] YAP T, PICEK S, and BHASIN S. Beyond the last layer: Deep feature loss functions in side-channel analysis[C]. Proceedings of the 2023 Workshop on Attacks and Solutions in Hardware Security, Copenhagen, Denmark, 2023: 73–82. doi: 10.1145/3605769.3623996.

[33] NI Lei, WANG Pengjun, ZHANG Yuejun, *et al*. Profiling side-channel attacks based on CNN model fusion[J]. *Microelectronics Journal*, 2023, 139: 105901. doi: 10.1016/j.mejo.2023.105901.

徐 杨：男，硕士生，研究方向为人工智能安全.

李锴彬：男，博士生，研究方向为深度学习，侧信道攻击，硬件安全.

何星星：男，副教授，研究方向为神经符号学习/推理.

Deep Side-Channel Attack Method Integrating Convolutional Block Attention Mechanism and Triplet Metric Learning

XU Yang[①] LI Kaibin[①] HE Xingxing[②]

[①](*School of Information Science and Technology, Southwest Jiaotong University, Chengdu* 611756, *China*)

[②](*School of Mathematics, Southwest Jiaotong University, Chengdu* 611756, *China*)

Abstract:

Objective Side-channel attack (SCA) constitutes one of the primary threats to the physical security of cryptographic chips, and leveraging deep learning techniques for key recovery has emerged as a research hotspot in the field of side-channel attack. However, existing deep analysis methods lack the ability to focus on key leakage intervals during feature extraction, especially in long-waveform and high-dimensional noise scenarios. They are easily disturbed by irrelevant background noise, leading to low feature extraction efficiency and slow guessing entropy convergence. To address these issues, this study proposes a deep side-channel analysis method integrating the Convolutional Block Attention Module (CBAM) and triplet loss, aiming to enhance the model's ability to capture weak leakage features in complex noise environments and improve key recovery efficiency.

Methods The proposed method introduces CBAM into the convolutional neural network (CNN) to construct an adaptive feature extraction network. CBAM includes two sub-modules: Channel Attention Module (CAM) and Spatial Attention Module (SAM). CAM adaptively adjusts the weights of different feature channels to emphasize channels containing high signal-to-noise ratio (SNR) leakage information, while SAM locates key Points of Interest (POI) in the time domain to suppress background noise in non-leakage intervals. After feature calibration by CBAM, triplet loss is adopted as the optimization objective to constrain the distribution of embedded features, forcing similar samples to form compact clusters and different samples to maintain sufficient separation in the feature space. Finally, a multivariate Gaussian template attack is implemented using the optimized embedded features to recover the key. The overall framework is shown in (Fig.2).

Results and Discussions Experiments are conducted on two public benchmark datasets (ASCAD and AES_HD) with guessing entropy (GE) and the minimum number of traces required for GE convergence to 1 (T_{GE0}) as evaluation metrics. On the ASCAD dataset: (1) In the ASCAD_f (HW) scenario, the proposed method only needs 144 traces to recover the key, reducing by 51.0% compared with the traditional CNN model; (2) In the ASCAD_f (ID) scenario, merely 61 traces are required, a reduction of 68.0% from the traditional

CNN; (3) In the random key setting (ASCAD_r), the method achieves convergence with 176 (HW) and 137 (ID) traces respectively, outperforming mainstream methods such as RL-SCA and Metric Learning (Table 2, Fig.3). On the low-SNR AES_HD dataset, the method reduces to 1 219 traces, which is lower than MHA and NLS, and the guessing entropy converges smoothly without obvious fluctuations (Table 2, Fig.4). Furthermore, the desynchronization experiments demonstrate that even under strong desynchronization noise interference, the proposed method can still adaptively lock onto effective leakage moments, exhibiting excellent robustness against time-domain jitter(Table 3). Ablation studies further confirm the positive synergistic effect of the proposed architecture and thoroughly validate the rationality of core components (Table 4).

Conclusions This study proposes an effective deep side-channel analysis method by integrating the CBAM attention mechanism and metric learning. The CBAM module enables the network to actively focus on key leakage information, solving the problem of insufficient focusing ability in traditional CNN-based methods. The triplet loss optimizes the discriminability of embedded features, further improving the accuracy of template matching. Experimental results on ASCAD and AES_HD datasets demonstrate that the method significantly reduces the number of traces required for key recovery and accelerates the convergence of guessing entropy, outperforming existing mainstream methods in both fixed and random key scenarios, as well as in low-SNR environments. Future work will focus on improving the adaptability of the model in scenarios with more severe synchronization disturbances, and enhancing its generalization ability in small sample scenarios.

Key words: Side-channel attacks; Deep learning; CBAM; Triplet loss; Hardware security