DOI:10.19651/j.cnki.emt.2517964

基于 SecureViT 的恶意代码检测模型*

张 傲 刘 薇 刘 阳 李 波 刘芳菲
(沈阳理工大学信息科学与工程学院 沈阳 110159)

摘 要:随着恶意代码的多样性和隐蔽性不断增加,传统的恶意代码检测方法在面对未知恶意代码时往往面临高成本和不稳定性的挑战。本研究旨在提出一种轻量化且高效的恶意代码检测模型,以适应资源受限环境中的应用需求。本文提出了一种基于 SecureViT 的轻量化恶意代码检测模型。该模型通过引入 ACF 模块与 MSDC 模块实现高效特征提取与精准分类。ACF 模块增强了模型对全局上下文信息的建模能力,MSDC 模块则通过多尺度特征提取与动态显著性调整进一步提升特征表达的丰富性。实验结果表明,SecureViT 模型在 Malimg、Virus-MNIST 和 BIG2015 数据集上的分类精度分别为 97.46%、91.17%和 95.49%,且计算开销仅为 1.71 GMAC,显著提高了检测性能并有效降低了计算成本。该模型在恶意代码检测中展现了优异的检测精度与低计算复杂度,具备在资源受限环境中的实际应用潜力。

关键词:恶意代码检测;上下文融合;多尺度卷积;轻量化深度学习模型
中图分类号:TN918.1 **文献标识码**:A **国家标准学科分类代码**:510.6120

Malicious code detection model based on SecureViT

Zhang Ao Liu Wei Liu Yang Li Bo Liu Fangfei
(College of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China)

Abstract: With the increasing diversity and concealment of malicious code, traditional detection methods often face high costs and instability when dealing with unknown malware. This study aims to propose a lightweight and efficient malware detection model to meet the application requirements in resource-constrained environments. This paper proposes a lightweight malware detection model based on SecureViT. The model achieves efficient feature extraction and accurate classification by introducing the ACF module and MSDC module. The ACF module enhances the model's ability to model global context information, while the MSDC module further improves the richness of feature representation through multi-scale feature extraction and dynamic significance adjustment. Experimental results show that the SecureViT model achieves classification accuracies of 97.46%, 91.17%, and 95.49% on the Malimg, Virus-MNIST, and BIG2015 datasets, respectively, with a computational cost of only 1.71 GMAC, significantly improving detection performance and effectively reducing computational costs. This model demonstrates excellent detection accuracy and low computational complexity, making it highly applicable in resource-constrained environments.

Keywords: malicious code detection; context fusion; multi-scale convolution; lightweight deep learning model

0 引 言

近年来,深度学习技术,尤其是卷积神经网络(convolutional neural network,CNN)和 Transformer 模型,凭借其强大的自动特征提取能力,逐渐在恶意代码检测领域得到了广泛应用。与传统依赖人工设计特征的方法相比,深度学习模型能够从原始恶意代码数据(如二进制文件或图像形式)中自动提取特征,大幅提升了恶意代码检测的性能。例如,文献[1]首次将预训练于 ImageNet 数据集的 VGG16 和 ResNet50 网络迁移至恶意代码图像分类任务,验证了迁移学习的有效性。文献[2]提出了集成 VGG16 和 ResNet50 的迁移微调策略 IMCEC,以提高特征提取质量和检测精度。与此同时,文献[3]和文献[4]分别采用了 AlexNet、ResNet152、DenseNet201 和 ShuffleNet 等网络架构进行集成优化,进一步提升了恶意代码分类的性能。文献[5]提出了一种通过引入轻量级卷积 Ghost 技术来实现

收稿日期:2025-01-19
*基金项目:辽宁省教育厅高等学校基本科研项目(JYTMS20230189)、沈阳理工大学引进高层次人才科研支持计划(1010147001131)项目资助

模型轻量化，从而有效减少计算开销的方法。

然而，尽管深度学习方法在恶意代码检测中取得了显著进展，现有方法仍存在一定的局限性。首先，CNN 虽然能有效提取局部特征，但在捕捉恶意代码的全局依赖关系方面存在不足。因此，Transformer 模型因其在长程依赖关系建模和全局信息处理方面的独特优势，逐渐被引入恶意代码检测领域。文献[6]提出的 Transformer 模型，通过自注意力机制，能够有效捕捉输入数据的长程依赖关系，显著提升了模型性能。然而，Transformer 模型虽然在全局建模方面有显著优势，但其计算复杂度较高，尤其在处理大规模数据时，训练和推理的效率成为了实际应用的瓶颈。

为了解决这些问题，近年来研究者尝试结合 CNN 和 Transformer 技术，设计混合模型，以兼顾高效性和强特征表达能力。例如，文献[7]提出的 ViT4Mal 模型将视觉 Transformer(vision transformer，ViT)与 CNN 相结合，用于边缘设备上的恶意软件检测。该模型通过 CNN 提取局部特征并将可执行代码转换为图像，再通过 ViT 捕获全局信息，从而在确保接近 97%的精度的同时，显著降低了计算复杂度。此外，文献[8]提出的基于 ViT 框架的自注意力机制方法，通过增强特征的可解释性和全局建模能力，在 Android 系统的恶意软件检测中取得了 80.27%的精度。文献[9]提出了将 ViT 与 CNN 结合的混合模型，在恶意软件图像分类任务中，通过结合 CNN 的像素强度特征与 ViT 的全局建模能力，取得了 96.62%的精度，并显著减少了推理时间。文献[10]提出了轻量级恶意代码检测框架 FasterMalViT，该模型融合了 CNN 和 ViT，显著提升了恶意代码检测的性能，尤其在处理大规模数据集时，大幅提高了检测准确性。

这些研究表明，将 CNN 的高效局部特征提取能力与 Transformer 的强大全局建模能力结合，能够有效平衡计算复杂度与检测性能。然而，尽管这一方向已取得显著进展，但目前的混合模型仍然面临一定的局限性。首先，虽然计算效率有所提升，但在大规模恶意代码检测任务中，模型的复杂度和计算需求依然较高，限制了其在实际应用中的部署。其次，现有模型在面对多样化和复杂的恶意代码时，特征提取的能力仍存在瓶颈，尤其是如何在保证精度的同时，降低计算成本，仍是一个亟待解决的问题。

为此，本文提出了一种新的恶意代码检测模型——SecureViT。该模型融合了自适应卷积融合(adaptive convolution fusion，ACF)模块和多尺度动态卷积(multi-scale dynamic convolution，MSDC)模块，旨在解决现有方法的局限性。ACF 模块通过结合静态上下文特征和动态注意力机制，显著增强了模型对恶意代码全局上下文信息的建模能力，同时有效优化了计算效率。MSDC 模块通过深度卷积(depthwise convolution，DWConv)在多个尺度上捕捉恶意代码的局部特征，从而进一步提升了检测性能和特征表达能力。该模型不仅在准确性上表现优秀，而且在计算复杂度上具有显著优势，尤其适用于资源受限的实际应用场景。

1　SecureViT 模型

SecureViT 是一种改进型混合架构，融合了 CNN 与 Transformer 技术，充分结合了 CNN 的空间归纳偏置能力与视觉 Transformer 的全局特征处理能力。在保留 MobileViT 模型轻量化优势的基础上，SecureViT 进一步通过 ACF 模块和 MSDC 模块，显著提升了特征表达能力与分类性能。其架构主要由 MV2 模块和改进后的 MobileViT block[11]组成，二者通过级联连接实现局部和全局特征的高效提取与融合，如图 1 所示。

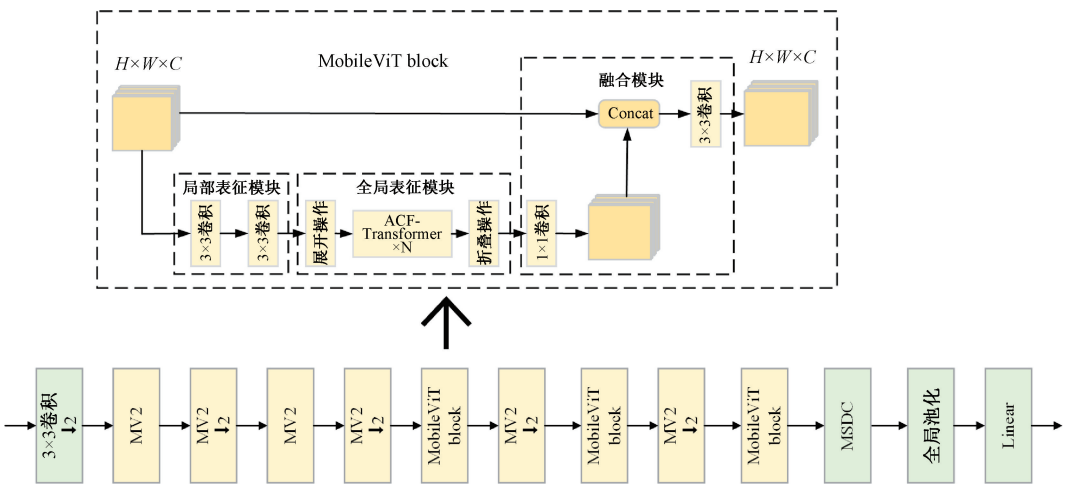

图 1　SecureViT 模块架构

Fig. 1　SecureViT module architecture

1.1　ACF-Transformer 模块

恶意代码的复杂性与多变性对检测模型提出了两大核心挑战:如何同时捕获局部与全局特征以提升分类精确率,以及如何在有限计算资源下实现高效推理。针对这些问题,本文提出了一种创新的 Transformer 架构——ACF-Transformer(如图 2 所示),其核心是通过 ACF 模块来增强模型对局部与全局信息的捕获能力,同时显著优化计算效率。与传统 Transformer 架构不同,ACF-Transformer 通过引入了静态上下文编码与动态上下文特征相结合的模块设计,以更好地适应多样化的任务需求。

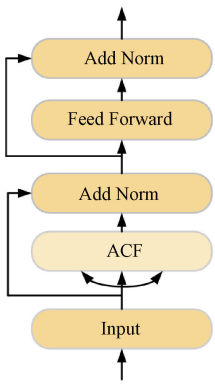

图 2　ACF-Transformer 结构

Fig. 2　ACF-Transformer structures

1)ACF 模块

恶意代码具有复杂多样的特征分布,包括局部行为模式和全局行为模式。传统方法往往侧重于局部特征提取(如基于卷积的模型),导致全局上下文信息不足;而引入自注意力机制的模型虽然增强了全局特征的捕捉能力,但却容易忽视局部上下文特性,影响特征的完整性和表达能力。ACF 模块通过联合静态上下文特征编码与动态注意力机制,显著增强了上下文特征表达能力,同时在计算效率上具有明显优势。ACF 模块结构如图 3 所示。

首先,对输入特征矩阵 $\boldsymbol{X} \in \mathbb{R}^{B\times C\times H\times W}$ 进行处理,提取静态特征 K_1 和动态特征 V,具体地,静态特征通过深度卷积提局取部,公式如下:

$$K_1 = \mathrm{Conv_d}(\boldsymbol{X}) \tag{1}$$

其中,$\mathrm{Conv_d}$ 表示深度卷积操作。

动态特征 V 使用标准 1×1 卷积操作生成,用于捕捉输入特征的多样性:

$$V = \mathrm{BN}(\mathrm{Conv_{1\times1}}(\boldsymbol{X})) \tag{2}$$

接下来,静态特征 K_1 与输入特征 $\boldsymbol{X}$ 在通道维度拼接形成组合特征 L,然后通过注意力嵌入模块进行特征权重计算:

$$L = [K_1, \boldsymbol{X}] \in \mathbb{R}^{B\times 2C\times H\times W} \tag{3}$$

利用两层卷积和激活函数计算特征融合的注意力权重 A:

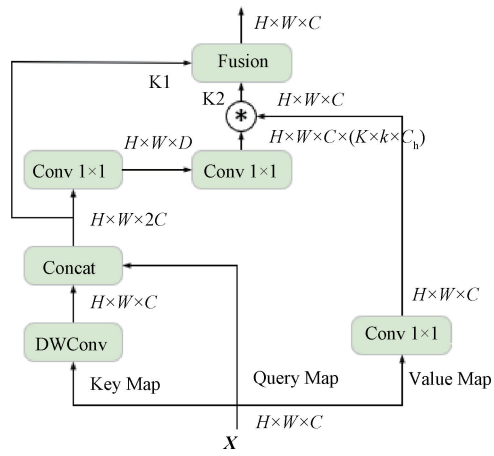

图 3　ACF 模块结构

Fig. 3　ACF module structure

$$A = \mathrm{Conv_{1\times1}}(\sigma((Conv_{1\times1}(L)))) A \in \mathbb{R}^{B\times(k\times k)\times C\times H\times W} \tag{4}$$

其中,$k\times k$ 表示卷积核的维度。随后,计算每个位置的平均权重:

$$\hat{A} = \frac{1}{k\times k}\sum_{i=1}^{k\times k} A_i \tag{5}$$

最后,利用 Softmax 函数对权重进行归一化以得到注意力权重矩阵:

$$\boldsymbol{\alpha}_{ij} = \frac{exp(\hat{A}_{ij})}{\sum_{j=1}^{HW} exp(\hat{A}_{ij})}, \boldsymbol{\alpha} \in \mathbb{R}^{B\times C\times H\times W} \tag{6}$$

在特征融合阶段,通过加权方式融合静态和动态特征,生成最终的输出特征矩阵 $\boldsymbol{Y}$。首先,动态上下文表示 K_2 通过权重与动态特征 V 相乘得到:

$$K_2 = \alpha \odot V,\ K_2 \in \mathbb{R}^{B\times C\times H\times W} \tag{7}$$

其中,$\odot$ 表示逐元素相乘。

最终输出特征 O 为静态特征 K_1 和动态特征 K_2 的加权和:

$$\boldsymbol{Y} = K_1 + K_2 \tag{8}$$

$$\boldsymbol{Y} = Y.reshape(B, C, H, W) \tag{9}$$

这种融合方式能够灵活地调整静态和动态特征的重要性。

1.2　MSDC 模块

传统的卷积操作通常在固定的尺度下提取特征,难以同时捕获多尺度信息。然而,恶意代码的行为模式和特征往往呈现多样性和多层次分布,仅靠单一尺度特征无法充分描述这些复杂性,导致分类性能受限。另外,以往的特征融合方法多采用简单的特征拼接或逐元素加法,未能针对输入特征的重要性进行动态调整。这样的方法可能导致关键特征表达被稀释,噪声特征占据较大比重,从而影响模型的整体性能。针对以上问题本文提出了一种名为 MSDC 的模块,其核心思想在于通过多尺度特征提取、动

态特征融合相结合，构建一个轻量化且高效的特征提取框架。该机制的结构如图 4 所示。

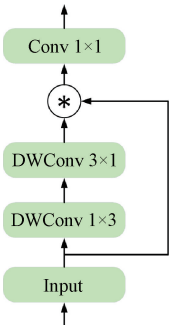

图 4　MSDC 模块结构

Fig. 4　Multi-scale dynamic convolution module structure

模块首先采用两个串联的深度卷积，即 $\mathrm{DWConv}_{1\times3}$ 和 $\mathrm{DWConv}_{3\times1}$，分别捕获输入特征 $X\in\mathbb{R}^{H\times W\times C}$ 的水平和垂直方向信息，从而提取多尺度特征[12]。其计算过程如下：

$$X_1=\mathrm{DWConv}_{1\times3}(X) \tag{10}$$

$$X_2=\mathrm{DWConv}_{3\times1}(X_1) \tag{11}$$

通过这种逐步分解的卷积操作，大幅度降低了计算复杂度，同时保证了空间特征的充分提取。在完成多尺度特征提取后，模块通过逐元素乘法（Element-wise Multiplication）实现动态特征融合。逐元素乘法是一种轻量化的特征融合方法，能够动态调整输入特征的显著性。具体地，融合后的特征表示为：

$$X_3=X\odot X_2 \tag{12}$$

其中，$\odot$ 表示逐元素乘法操作，通过该过程可以选择性地增强关键特征，抑制噪声和无关信息。

随后，为进一步整合通道信息并压缩计算量，模块采用 $\mathrm{Conv}_{1\times1}$ 对融合后的特征 X_3 进行通道维度的重构，最终生成输出特征 Y：

$$Y=\mathrm{Conv}_{1\times1}(X_3) \tag{13}$$

2　实验与结果分析

2.1　实验数据集

本实验基于 Malimg、Virus-MNIST 和 BIG2015 数据集进行评估，用于测试 SecureViT 模型在恶意代码图像分类任务中的性能。Malimg 数据集包含 9 339 张恶意代码家族图像，覆盖 25 个类别；Virus-MNIST 数据集包含 47 778 张样本，分为 10 类，类别分布相对均衡；而 BIG2015 数据集来源于真实场景，为在实现更均衡的数据分布，本文去除了图像数量较少的类别，并且从图像数量较多的类别中去除了一部分样本。图 5～7 分别展示了 Malimg 数据集、Virus-MNIST 数据集以及经过数据均衡处理后的 BIG2015 数据集中样本类别的具体分布情况。

2.2　评估指标

在本次实验中，为了全面评估 SecureViT 模型在恶意代码图像分类任务中的性能，采用了精确率（Precision）、召回率（Recall）、F1 分数（F1 Score）、模型参量和 GMAC 五个关键评价指标。

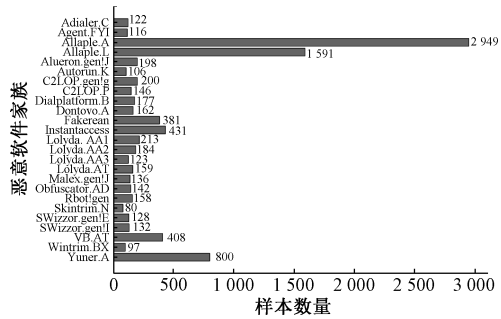

图 5　Malimg 数据集样本类别分布

Fig. 5　Sample category distribution for the Malimg dataset

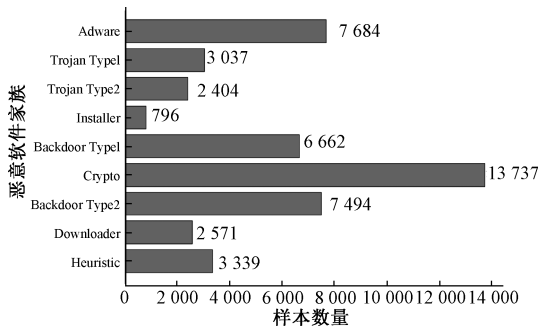

图 6　Virus-MNIST 数据集样本类别分布

Fig. 6　Sample category distribution for the Virus-MNIST dataset

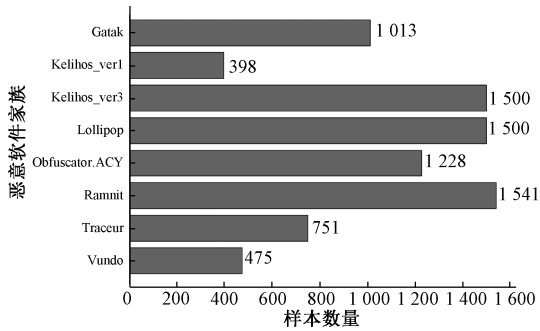

图 7　BIG2015 数据集样本类别分布

Fig. 7　Sample category distribution for the BIG2015 dataset

2.3　实验环境和超参数设置

本研究实验环境基于 Linux（Ubuntu 20.04）操作系统，使用 Python 3.8 编程语言和 PyTorc 1.11.0 深度学习框架完成模型的开发与训练。硬件配置方面，实验平台搭载 NVIDIA RTX 4090D 图形处理器。此外，实验环境支持 CUDA 11.3 技术，在模型训练过程中充分发挥 GPU 加速的优势，大幅提升了计算效率与模型性能。在训练过程的超参数设置中，输入尺寸为 224 × 224，优化器为 AdamW，学习率为 0.000 2，批处理大小为 128，权重衰减为 0.001。

2.4　模型性能与分析

为验证本文提出的 SecureViT 模型在恶意代码检测任务中的有效性与鲁棒性,实验在 Malimg 数据集上进行评估。重点分析模型的收敛性和分类能力。

从图 8(验证精度曲线)可见,模型在训练的初始阶段精度显著提升,体现了模型较强的特征学习能力。随后,随着训练的进行,验证精度趋于稳定并在后期保持小幅波动,最终达到 97.46%的峰值。这表明模型在特征提取与优化过程中具有良好的全局收敛性,同时有效避免了过拟合现象。图 9(验证损失曲线)反映了模型在训练过程中的优化效果。初始损失值较高,表明模型初期尚未捕获有效的特征。然而,通过多轮训练,损失值迅速下降并在约 10 轮时趋于平稳,最终收敛至 0.2 左右。这一结果验证了 SecureViT 模型在特征学习与参数优化上的高效性,能够在较短时间内完成有效的特征建模与分类。图 10(混淆矩阵)进一步揭示了模型在多类别分类任务中的性能表现。可以观察到,模型对绝大多数类别的恶意代码样本均实现了高精度分类,仅少量样本出现错误分类,且错误分布较为随机,未表现出特定类别间的显著混淆。这反映了模型较强的泛化能力与鲁棒性,尤其是在多类别任务中的适应性。

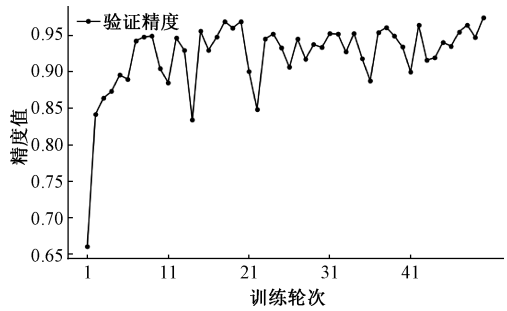

图 8　验证精度曲线

Fig. 8　Validation precision curve

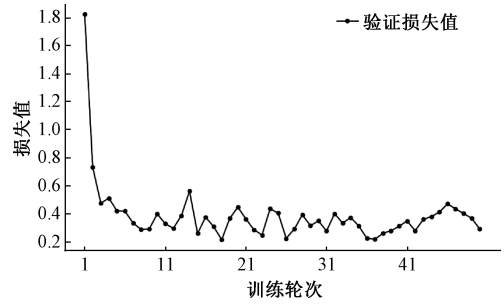

图 9　验证损失曲线

Fig. 9　Validation loss curve

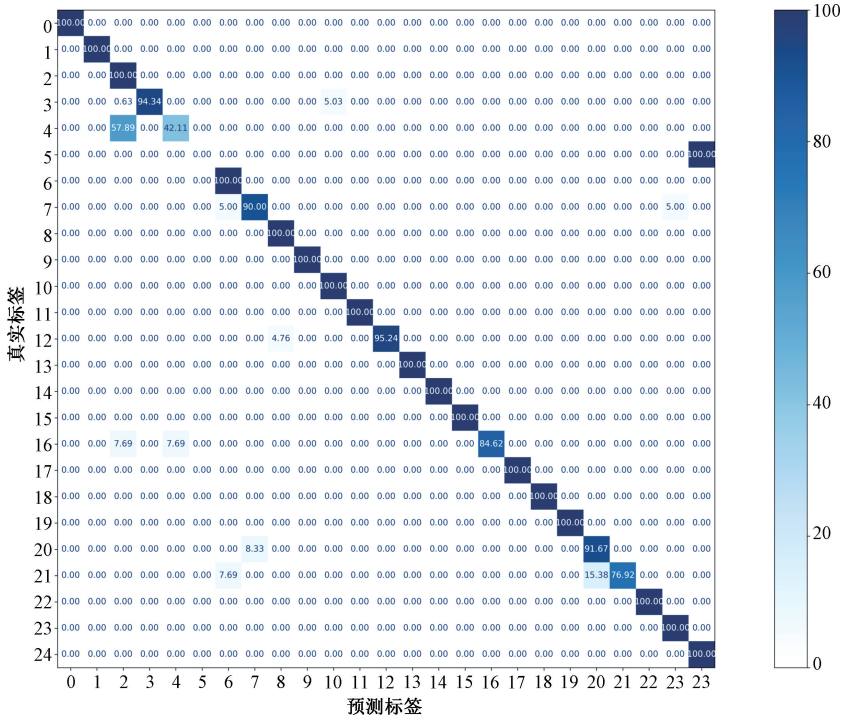

图 10　Malimg 数据集混淆矩阵

Fig. 10　Confusion matrix of Malimg datase

综上所述,本文提出的 SecureViT 模型在 Malimg 数据集上的分类性能表现卓越,精确率达到 97.46%,召回率为 98.37%,F1 分数为 97.91%。该模型在特征提取精度和全局建模能力上均展现出显著优势。通过 ACF 模块和 MSDC 模块的有机结合,SecureViT 在全局探索与局部优化之间达到了良好的平衡,成功实现了高效且精准的恶意

代码检测。

为了进一步验证本文提出的基于 SecureViT 模型在恶意代码检测任务中的性能，本文对比了多个主流深度学习模型，包括 ShuffleNet[13]、MobileNet[14]、GoogleNet[15]、ResNet[16]、ConvNeXt[17]、DenseNet[18]、VGG[19]、Swin transformer[20]、Vision Transformer[21]、Mobile ViT，以及其他恶意代码检测模型 3 D-VGG-16[22]、SPP CNN[23]、Xception-SE[24]、GLCM＋SVM[25]，表 1 中展示了各模型的 Precision、Recall、F1 分数以及模型参量和 GMAC。

从表 1 所示的实验结果可知，不同模型在恶意代码检测任务中的性能表现存在显著差异。轻量化模型，如 MobileNetV2 和 Shufflenet，尽管在计算复杂度和参数量上表现出较低的数值，但由于其特征提取能力相对有限，在处理复杂的恶意代码样本时，往往无法提供足够的精度和召回率。例如，MobileNetV2 的精度为 89.94%，召回率为 91.98%，均低于多种高效模型的表现。这一结果表明，单纯依赖轻量化设计在恶意代码检测任务中难以满足高精度的需求，尤其是在面对具有高度变异性和复杂性的恶意代码时，模型需要具备更为强大的特征提取和识别能力。

表 1　对比实验结果

Table 1　Compares the experimental results

模型	Precison	Recall	F1	模型参量/M	GMAC
EfficientNet-B0	94.70	94.90	94.80	4.04	0.406
EfficientNet-B1	93.51	93.62	93.56	6.55	0.599
EfficientNet-B2	94.55	95.15	94.85	7.74	0.689
EfficientNet-B3	96.00	96.40	96.20	10.73	1.0
MobileNetV2	89.94	91.98	90.95	2.26	319.11
MobileNetV3-Large	91.20	89.27	90.22	4.23	229.47
MobileNetV3-Small	86.51	85.80	86.15	1.54	59.62
Shufflenet V2 ×0.5	91.50	93.71	92.59	0.369	0.043
Shufflenet V2 ×1.0	96.15	96.92	96.53	1.28	0.151
GoogLeNet	80.87	79.30	80.08	10.01	1.51
ResNet-34	96.19	97.08	96.63	21.29	3.68
ResNet-50	96.33	97.18	96.75	23.56	4.13
ResNet-101	97.18	96.30	76.74	42.55	7.86
ConvNeXt-Small	96.25	97.15	96.70	49.47	8.73
ConvNeXt-Base	96.81	97.82	97.31	87.59	15.42
DenseNet-161	97.02	97.94	94.48	26.53	7.85
VGG-11	85.09	79.63	82.27	128.87	7.64
VGG-16	79.40	74.65	96.95	134.36	15.52
VGG-19	77.75	72.91	75.25	139.67	19.68
Swin Transformer-Tiny	69.97	78.38	73.94	27.54	4.38
Transformer-Base	65.34	63.38	64.35	85.82	16.88
Mobile ViT	95.58	96.42	96.00	4.96	1.56
3D-VGG-16	97.00	96.00	96.50	—	—
SPP CNN	96.80	97.10	96.95	—	—
Xception-SE	92.20	91.83	92.02	—	—
GLCM＋SVM	93.40	93.00	93.20	—	—
本文模型	97.46	98.37	97.91	6.26	1.71

经典的深度卷积神经网络（如 ResNet 和 DenseNet）通过增加网络的深度以及引入多尺度特征提取机制，显著提高了恶意代码特征的捕捉能力。在本实验中，ResNet-50 的精度达到了 96.33%，召回率为 97.18%，相较于轻量化模型，其性能优势较为明显。然而，这类网络由于参数量庞大及计算开销较高，限制了其在资源受限环境中的应用，尤其是在移动设备或实时检测系统中，较高的计算复杂度使得这些网络难以发挥其性能优势，导致其在实际应用中的可行性受到一定制约。

相比之下，基于 Transformer 架构的模型因其出色的

全局特征提取能力,已在视觉任务中展现出了巨大的潜力。例如,Mobile ViT 作为一种轻量化的 Transformer 模型,凭借其强大的全局上下文建模能力,在恶意代码检测任务中取得了 95.58%的精度和 96.42%的召回率。尽管该模型在捕捉全局特征方面相比传统卷积神经网络具有一定优势,但其模型精度仍有提升空间。

最终,本文提出的 SecureViT 模型在分类精度(97.46%)和召回率(98.37%)上显著优于大多数对比模型,同时在计算复杂度方面表现卓越,计算开销仅为 1.71 GMAC。这一性能表明,SecureViT 模型在保证高检测精度的同时,能够大幅降低计算需求,从而在高效性和准确性之间达到了优良的平衡。该模型在资源受限的应用场景下,尤其是在实时恶意代码检测中,展现出了极大的应用潜力和实践价值。

2.5　消融实验

为了评估 SecureViT 模型中不同模块对整体分类性能的影响,本研究设计了 7 组对比实验,通过逐步移除或替换模型的关键模块,来验证各个模块对模型表现的贡献。实验结果展示了不同模块的去除或替换对模型参数、计算复杂度以及 F1 分数的影响。

如表 2 所示,消融实验结果表明,基线模型 Mobile ViT 的 F1 分数为 96.00%,计算复杂度为 1.56 GMAC,展示了其作为轻量化模型的性能平衡性,但在捕捉复杂恶意代码特征方面存在局限性。加入 MSDC 后,F1 分数提升至 96.75%,证明其在多尺度特征提取和上下文信息建模方面的显著作用,且未显著增加模型参数量和计算复杂度。而加入 ACF 后,尽管 F1 分数达到 96.64%,表现略低于 MSDC,但其计算复杂度增加至 1.71 GMAC,表明该模块对资源的消耗较高。进一步加入 EMA Attention 和 CoT Attention 后,分别将 F1 分数提升至 96.80%和 97.38%。加入 Additive Attention 虽将 F1 分数提升至 96.60%,接近 ACF 的表现,但其计算复杂度降至 1.45 GMAC,展现了作为轻量化替代方案的高效性。最终,本文模型结合 ACF 和 MSDC 的设计,实现了 97.91%的 F1 分数,显著优于其他对比模型。尽管参数量和计算复杂度略高于基线模型,但其显著的性能提升证明了本文提出模块设计的有效性,并在捕捉复杂特征的同时实现了性能与计算效率的良好平衡,为恶意代码检测任务提供了高效的解决方案。

表 2　消融实验结果

Table 2　Results of the ablation experiments

模型	Precison	Recall	F1	模型参量/M	GMAC
Mobile ViT	95.58	96.42	96.00	4.96	1.56
Mobile ViT+MSDC	96.34	97.18	96.75	4.96	1.56
Mobile ViT+ACF	96.24	97.05	96.64	6.25	1.71
Mobile ViT+EMA Attention	96.42	97.18	96.80	6.25	1.71
Mobile ViT+CoT Attention	96.94	97.83	97.38	6.26	1.71
Mobile ViT+Additive Attention	96.24	96.97	96.60	4.96	1.45
本文模型	97.46	98.37	97.91	6.26	1.71

2.6　正则化方法对模型性能的影响

在深度学习模型中,正则化方法通过抑制过拟合,提升模型的泛化能力。为了研究正则化对 SecureViT 模型在恶意代码检测任务中的性能影响。表 3 实验结果表明,正则化方法对 SecureViT 模型的性能具有显著影响。其中,L1 正则化通过对参数稀疏化的约束,显著抑制了过拟合,取得了最佳的分类性能,F1 分数达到 97.91%,Precision 和 Recall 分别为 97.46%和 98.37%;相比之下,L2 正则化的 F1 分数为 96.84%,在抑制过拟合和提升泛化能力方面效果略逊于 L1 正则化。而在未使用正则化时,模型仍然保持了较高的性能,F1 分数为 97.39%,但模型的泛化能力较低,存在过拟合风险。综上,选择合适的正则化方法,特别是 L1 正则化,是提升模型在恶意代码检测任务中鲁棒性和泛化能力的关键。

表 3　正则化方法对模型性能的影响

Table 3　Effect of the regularization method on the model performance

Regularization Methods	Precision	Recall	F1
L1	97.46	98.37	97.91
L2	96.14	97.56	96.84
无正则化	97.04	97.75	97.39

2.7　泛化实验

为评估所提出的 SecureViT 模型的泛化能力和稳定性,本文进行了 5 次独立实验,选取 Virus-MNIST 和 BIG2015 数据集作为测试数据集。实验结果(如表 4 所示)显示,SecureViT 在这两个数据集上均表现出较强的泛化能力,并与其他主流模型进行了对比。

在 Virus-MNIST 数据集上,SecureViT 模型的 Precision 平均值为 91.17%,方差为 0.39,优于 ResNet-50(Precision 平均值为 89.49%,方差为 0.51)和 Mobile ViT(Precision 平均值为 89.34%,方差为 0.45)。这一结果表

表 4　泛化实验结果

Table 4　Results of the generalization experiments

数据集	Model	Precision 平均值	Precision 方差
Virus-MNIST	ResNet-50	89.49	0.51
	Mobile ViT	89.34	0.45
	本文模型	91.17	0.39
BIG2015	ResNet-50	93.80	0.42
	Mobile ViT	93.53	0.34
	本文模型	95.49	0.37

明，SecureViT 在该数据集上不仅在准确度上超越了其他模型，还在稳定性上具有优势。在 BIG2015 数据集上，SecureViT 的 Precision 平均值为 95.49%，方差为 0.37，优于 ResNet-50（Precision 平均值为 93.80%，方差为 0.42）和 Mobile ViT（Precision 平均值为 93.53%，方差为 0.34）。尽管 Mobile ViT 在该数据集上表现出更低的方差，但 SecureViT 依然以较高的 Precision 平均值和较低的 Precision 方差表现出较好的稳定性，凸显了其卓越的泛化能力和稳定性。

3　结　　论

本文提出了一种创新的基于 SecureViT 的轻量化恶意代码检测模型，通过引入 ACF 模块和 MSDC 模块，显著提升了检测精度与效率。模型的核心创新在于结合动态与静态上下文信息，增强了模型对复杂恶意代码样本的适应性，并通过多尺度特征提取与动态显著性调整，强化了全局特征的建模能力。

在 Malimg、Virus-MNIST 和 BIG2015 数据集的对比实验中，所提模型分别在这 3 个数据集上取得了 97.46%、91.17%和 95.49%的分类精度，并且在计算开销方面保持了较低水平，展示了其在资源受限环境中的实用性。经进一步的实验与分析，该模型在恶意代码检测任务中表现出了较强的全局搜索能力与局部优化能力，有效避免了过拟合问题。

未来的工作将着眼于进一步提升该模型在处理未知恶意代码上的泛化能力。计划将该模型应用于更广泛的恶意代码类型检测，探索其与其他深度学习模型的融合，进一步提升模型的检测精度与效率。此外，针对实际应用环境中的资源限制问题，还将优化算法的计算复杂度，使其能够在更多嵌入式系统和移动端设备中高效运行。

参考文献

[1] REZENDE E, RUPPERT G, CARVALHO T, et al. Malicious software classification using VGG16 deep neural network's bottleneck features[C]. Information Technology-New Generations: 15th International Conference on Information Technology. Springer International Publishing, 2018:51-59.

[2] VASAN D, ALAZAB M, WASSAN S, et al. Image-Based malware classification using ensemble of CNN architectures (IMCEC)[J]. Computers & Security, 2020, 92: 101748.

[3] ASLAN Ö, YILMAZ A A. A new malware classification framework based on deep learning algorithms[J]. IEEE Access, 2021, 9: 87936-87951.

[4] WONG W K, JUWONO F H, APRIONO C. Vision-based malware detection: A transfer learning approach using optimal ECOC-SVM configuration[J]. IEEE Access, 2021, 9: 159262-159270.

[5] 李怡，李进. 基于 Ghost-DenseNet-SE 的恶意代码检测方法[J]. 空军工程大学学报(自然科学版)，2021，22(5):49-55.
LI Y, LI J. Malicious code detection method based on Ghost-DenseNet-SE [J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(5): 49-55.

[6] VASWANI A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30, DOI: 10.1109/NIPS2017-3.0006.

[7] RAVI A, CHATURVEDI V, SHAFIQUE M. ViT4Mal: Lightweight vision transformer for malware detection on edge devices[J]. ACM Transactions on Embedded Computing Systems, 2023, 22 (S5): 10.1145/3609112.

[8] JO J, CHO J, MOON J. A malware detection and extraction method for the related information using the ViT attention mechanism on android operating system[J]. Applied Sciences, 2023, 13(11): 6839.

[9] ASHAWA M, OWOH N, HOSSEINZADEH S, et al. Enhanced image-based malware classification using transformer-based convolutional neural networks (CNNs)[J]. Electronics, 2024, 13(20): 4081.

[10] 黄保华，杨婵娟，熊宇，等. 基于 ViT 的轻量级恶意代码检测架构[J]. 信息网络安全，2024，24(9):1409-1421.
HUANG B H, YANG CH J, XIONG Y, et al. Lightweight malicious code detection architecture based on ViT[J]. Information Network Security, 2024, 24(9): 1409-1421.

[11] MEHTA S, RASTEGARI M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer[J]. ArXiv preprint arXiv:2110.02178, 2021.

[12] 赵亚凤，宋文华，刘晓璐，等. 基于改进 YOLOv7 的钢轨缺陷检测方法[J]. 电子测量技术，2024，47(20):177-185.

ZHAO Y F, SONG W H, LIU X L, et al. Steel rail defect detection method based on improved YOLOv7[J]. Electronic Measurement Technology, 2024, 47(20): 177-185.

[13] ZHANG X Y, ZHOU X Y, LIN M X, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.

[14] WANG W, LI Y T, ZOU T, et al. A novel image classification approach via dense-MobileNet models[J]. Mobile Information Systems, 2020, 2020(1): 7602384.

[15] TANG P J, WANG H L, KWONG S. G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition[J]. Neurocomputing, 2017, 225: 188-197.

[16] TARG S, ALMEIDA D, LYMAN K. Resnet in resnet: Generalizing residual architectures[J]. ArXiv preprint arXiv:1603.08029, 2016.

[17] BENCHALLAL F, HAFIANE A, RAGOT N, et al. ConvNeXt based semi-supervised approach with consistency regularization for weeds classification[J]. Expert Systems with Applications, 2024, 239: 122222.

[18] ZHU Y, NEWSAM S. Densenet for dense flow[C]. 2017 IEEE International Conference on Image Processing(ICIP). IEEE, 2017: 790-794.

[19] SENGUPTA A, YE Y T, WANG R, et al. Going deeper in spiking neural networks: VGG and residual architectures[J]. Frontiers in Neuroscience, 2019, 13: 95.

[20] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[21] HAN K, WANG Y H, CHEN H T, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 87-110.

[22] AL-KHATER W, AL-MAEED S. Using 3D-VGG-16 and 3D-Resnet-18 deep learning models and FABEMD techniques in the detection of malware[J]. Alexandria Engineering Journal, 2024, 89: 39-52.

[23] 刘薇. 基于卷积神经网络的恶意代码灰度图像分类研究[D]. 北京:北京交通大学,2021:17-38.
LIU W. Research on malware grayscale image classification based on convolutional neural networks[D]. Beijing: Beijing Jiaotong University, 2021: 17-38.

[24] 蒋瑞林, 覃仁超. 基于深度可分离卷积的多神经网络恶意代码检测模型[J]. 计算机应用, 2023, 43(5): 1527-1533.
JIANG R L, QIN R CH. Malware detection model based on deep separable convolutions and multi-neural networks[J]. Computer Applications, 2023, 43(5): 1527-1533.

[25] CUI ZH H, XUE F, CAI X J, et al. Detection of malicious code variants based on deep learning[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3187-3196.

作者简介

张傲,硕士研究生,主要研究方向为数据安全与数据处理。

E-mail:1797718602@qq.com

刘薇(通信作者),博士,教授,主要研究方向为智能优化算法及其应用。

E-mail:LiuWei19781020@126.com