基于改进CNN的恶意软件分类方法

轩勃娜,李 进
(空军工程大学防空反导学院,陕西西安 710051)

摘 要: 越来越多的恶意软件变种给网络安全带来了巨大的威胁,导致了现有基于CNN(Convolutional Neural Networks)的恶意软件分类方法的泛化能力弱和准确性不足.为了解决这些问题,本文提出了一种新的方法,即基于改进CNN的恶意软件RGB(Red Green Blue)可视化的分类方法,可以抵御变种和混淆性恶意软件.首先,提出了一种基于RGB图像的特征表示方法,该方法更加关注恶意软件的二进制和汇编信息、API信息间的语义关系,生成具有更丰富纹理信息的图像,可以挖掘恶意代码原始与变种之间更深层的依赖关系.其次,针对恶意软件的加密和混淆问题,使用坐标注意力模块(Coordinate Attention Module,CAM)获取更大范围的空间信息来强化特征.最后,结合空洞空间金字塔池化(Atrous Spatial Pyramid Pooling,ASPP)来改进CNN模型,解决因图像尺寸归一化导致的信息丢失和冗余.实验结果表明,上述方法在最近的先进方法中脱颖而出,对Kaggle数据集和DataCon数据集的准确率分别达到99.48%和97.78%.与其它方法相比,该方法对Kaggle数据集的准确率提高了0.22%,对DataCon数据集的准确率提高了0.80%.本文方法可以有效地分类恶意软件和恶意软件家族变种,具有良好的泛化能力和抗混淆能力.

关键词: 网络安全;恶意代码分类;RGB图像;汇编信息;语义关系;坐标注意力模块;空洞空间金字塔
基金项目: 国家自然科学基金(No.61806219,No.61703426,No.61876189)
中图分类号: TP309.5 **文献标识码:** A **文章编号:** 0372-2112(2023)05-1187-11
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20220818

Malware Classification Method Based on Improved CNN

XUAN Bo-na, LI Jin
(*School of Air and Missile Defense, College of Air Force Engineering University of China, Xi'an, Shaanxi* 710051, *China*)

Abstract: The increasing variants malware bring a great threat to network security, leading to weak generalization and insufficient accuracy of existing base on the convolutional neural networks (CNN) malware classification methods. To solve these problems, an approach, namely, a classification method based on improved the CNN for malware RGB (Red Green Blue) visualization that can resist variants and obfuscation malware. Firstly, our method proposed a feature representation method based on RGB image, which pays more attention to the semantic relationship between binary, assembly information and API information of malware. The generated image, with richer vein information, that can uncover deeper dependencies between the original and variants of the malware. Secondly, to address the problems of malware encryption and obfuscation, this paper uses the coordinate attention module (CAM) to obtain a larger range of the spatial information to strengthen malware features. Finally, the Atrous spatial pyramid pooling (ASPP) is combined to improve the CNN model to address the information loss and redundancy due to image size normalization. The experimental results show that the above methods stands out among the recent advanced methods with an accuracy of 99.48% and 97.78% for dataset Kaggle and dataset DataCon. Compared with the other methods, our method had the accuracy increased by 0.22% for dataset Kaggle, and had the accuracy increased by 0.80% for dataset DataCon. Our method can effectively classify malware and variants of malware families, which has excellent generalization ability and anti-obfuscation ability.

Key words: network security; malware classification; RGB image; compile information; semantic relationship; coordinate attention module; atrous spatial pyramid pooling
Foundation Item(s): National Natural Science Foundation of China (No.61806219, No.61703426, No.61876189)

收稿日期:2022-07-14;修回日期:2022-11-02;责任编辑:覃怀银

1 引言

据估计,全球恶意软件产业价值数百万甚至数十亿美元,并且每年都在持续增长[1]. 著名的互联网安全公司赛门铁克,在其2018年互联网安全威胁报告(Internet Security Threat Report)[2]中指出从“WannaCry和Petya/NotPetya”的突然传播开始,产生了68个恶意家族和超过10 000种恶意变种代码. 恶意软件变种的数量呈指数级增长,使用传统的防病毒产品单独确定每个样本的恶意行为已经成为一个挑战[3,4].

随着人工智能技术的发展,结合恶意样本可视化和深度学习的检测和分类方法备受关注[5~8]. Cui[9]提出了一种基于深度学习方法的变种检测算法,首先使用蝙蝠算法来均衡数据,然后利用CNN(Convolutional Neural Networks)的模型来检测恶意软件变种. Fu等人[10]将恶意软件可视化为RGB(Red Green Blue)彩色图像并提取来自它们的全局纹理和颜色特征. 文献[11]提出了一种未使用逆向工程,仅从9个恶意软件系列中获取10 860个样本,达到了98.2%的分类准确度,但未考虑原始二进制文件中二进制代码的语义特征.

此外,使用深度学习模型[12~14]对于大规模恶意软件检测来说非常耗时,需要很多训练样本和较长的训练周期. Yuan等人[15]采用了不同的方式,提出了一种基于字节级恶意软件的解决方案,通过深度卷积VGG16模型,利用转移概率矩阵将二进制文件转换为马尔可夫图像,在Microsoft Malware数据集上获得了99.26%准确率. Qiao等人[16]提出一个基于多通道特征矩阵为恶意软件分类的LeNet5结构,使用Word2Vec技术将恶意软件二进制文件和汇编文件转换成特征矩阵.

总之,以上基于深度学习的方法大多是将恶意软件二进制文件转换成灰度图像或者图像特征,缺乏对特征间的语义相关性处理. 用深度神经网络训练时,灰度图像必须归一化处理,导致数据转换过程中产生冗余或者丢失信息的问题. 除此之外,传统方法直接检测效果不好,需要脱壳或者解密才可以识别,很难找到一种有效的解压缩和解密方法.

为提高对恶意代码变种的分类准确度和泛化能力,本文试图设计一种能够获取恶意代码特征间语义关系的可视化方法,并避免因图像归一化导致分类准确率低的问题. 同时,针对本文所使用的特征较为复杂多样的特点,设计实现基于改进CNN的恶意软件的RGB[17]可视化的分类方法,对恶意代码变种进行分类,并且尝试利用现网捕获的恶意代码对其进行验证.

2 反汇编可视化

在本节中,提出了一种将恶意软件转换为RGB图像的特征提取方法. 将二进制文件特征、汇编指令与数据和API(Application Programming Interface)信息结合在一起生成RGB图像,比单独利用来自某种特征信息更能反映恶意代码的空间特征. 生成的图像将具有更丰富的纹理信息,能够更好地反映恶意代码中样本之间的相关性.

2.1 二进制恶意软件可视化

本文采用了直接保留二进制代码所反映的信息. 恶意软件的二进制比特流可以被分成8比特的组,每个字节可以被视为一个像素. 像素值可以计算如下:

$$p_{\text{bin}}=b_7\times2^7+b_6\times2^6+b_5\times2^5+b_4\times2^4+b_3\times2^3+b_2\times2^2+b_1\times2^1+b_0\times2^0 \tag{1}$$

$$h_{\text{hex}}=h_1\times16^1+h_0\times16^0 \tag{2}$$

这里是$(b_7,b_6,b_5,b_4,b_3,b_2,b_1,b_0)$二进制的序列,$(h_1,h_0)$是十六进制的序列. 由于不同恶意软件家族的文件大小明显不同,生成为相同大小的图像是不合理的. 生成的灰度图像的大小,取决于虚拟地址数量,利用虚拟地址决定文件的长度,宽度利用文件大小进行计算,保证RGB三部分通过虚拟地址进行一一对应.

2.2 汇编指令和数据可视化

汇编指令和数据程序开发和编译的过程取决于各种因素,实现相同恶意功能的基本块并不一定由完全相同的操作码序列构成,导致实现相似功能的同族恶意代码所生成的图像矩阵并不一定相似.

2.2.1 提取操作码和操作数序列

汇编信息由两部分组成:操作码和操作数. 操作码可以反映程序的行为,所以同一家族的恶意软件在某种程度上会有相似的操作码序列,而操作数可以根据寄存器的类型进行标记. 因此,提取相应位置的操作码和操作数,并用0填充剩余位置.

2.2.2 对操作码和操作数进行编码

Kaggle数据集中收集1 473种不同类型的操作码,常用的235个类型出现比例达到99.97%,余下的1 218种类型仅占0.03%. 因此考虑采用常用的235种类型,并将其余的类型全部划分到第256种类型中. 为了有效归纳操作码,相同类型的操作码,进行相同的编码. 例如,将MOV、MOVQ和MOVD合并为1种操作码,都用于一般的数据传输功能. 经过筛选合并功能相同的操作码最终将操作码范围固定在235个,编码范围为1~235.

同理DataCon数据集经过反汇编、提取操作码和合并归纳操作后将操作码范围固定在235个,编码范围为1~235. 由于本文提出的编码范围小于灰度级像素值范围0~235,因此操作码的编码结果可以直接作为生成图像的像素值.

因此操作码和操作数的编码结果可以直接作为生成单通道图像G的像素值. 在提取的运算符和操作数显示在表1的第二行中,编码结果对应于第一行所示的虚拟地址,编码结果显示在第三行.

表1　Opcode编码实例

虚拟地址	10001000	10001001	10001002	10001003	10001004	10001005	10001006	10001007
操作码与操作数	MOV	Eax	Esp	Null	MOV	Eix	Eax	MOV
编码	4	235	0	0	4	0	235	4
虚拟地址	10001008	10001009	1000100A	1000100B	1000100C	1000100D	1000100E	1000100F
操作码与操作数	Edx	Eax	CMP	Eax	JNZ	Edx	MOV	Edx
编码	239	235	8	235	11	239	4	239

2.3　API可视化

在本文中,介绍了一种基于上下文相似性将相关API聚类分组到单个聚类名称中的方法. 将多个API分组到有限数量的动态库中,使得描述恶意代码的行为属性成为可能. 文献[18]中研究了API调用之间的上下文关系,以区分恶意行为. 然而,分类技术依赖于API序列模式,通过插入干扰信息很容易误导静态分析.

将从Kaggle数据集中,. asm反汇编文件中提取出的API信息和动态库信息. 而DataCon数据集利用IDA Pro工具反汇编成asm文件,从获取其中的API和动态库信息. 按照每个动态库内部顺序和外部调用关系,生成整个API调用关系序列.

在文件中通常表示为字母(A~Z、a~z)、数字(0~9)和符号(!、#、%)组合表示,其ASCII值在32到126之间. 在使用卷积神经网络中的卷积核提取局部特征时,其余字符加入会引起噪声. 编码范围小于灰度级像素值范围0~255,因此API的编码结果可以直接作为生成单通道图像B的像素值,具体如表2所示.

表2　API编码图解

虚拟地址	API接口	十六进制编码
00402000	GlobalFindAtom	47 6C 6F 62 61 6C 46 69 6E 64 41 74 6F 63 00 00
00402004	IsDBCSLeadByte	6C 74 44 42 43 53 4C 65 61 64 42 79 74 65 00 00
00402008	GetConsoleCP	47 65 74 43 6F 42 74 6C 65 43 50 00 00 00 00 00
0040200C	VirtualAlloc	56 69 72 74 75 61 6C 41 6C 6C 6F 63 00 00 00 00
00402010	CreateThread	43 72 65 61 74 65 54 68 72 65 61 64 00 00 00 00

3　框架概述

在数据预处理模块中,将恶意软件反汇编成RGB图像. 坐标注意力模块充分利用了注意力机制中的上下文学习能力,在训练过程中对更大范围的像素进行整合,更好获得全局感受野,并精确编码特征图的位置信息. 卷积神经网络采用非均匀的缩放策略来缩放模型,改进的渐进学习方法,利用深层卷积获取高维特征,然后提取的特征输入到空洞空间金字塔模块. 最后,在训练过程中调整优化器和损失函数. 如图1所示.

3.1　数据预处理模块

随着恶意代码对抗技术不断发展,导致恶意代码的变种呈现出多样化和动态化,针对恶意软件变种的分类需要更全面的特征,这是对恶意软件变种分类面临的新挑战. 在数据预处理模块中,用IDA Pro对恶意软件进行反汇编,然后按照虚拟地址提取恶意软件的静态特征,从而生成了RGB图像. 生成的图像大小取决于恶意样本内虚拟地址的数量,也就是恶意样本的大小. 本文属于预处理过程,如图2所示.

将二进制文件特征、汇编指令和数据和API聚类特征结合生成RGB图像,包含来自不同视角的信息,比单独利用来自这些视角的信息更能反映恶意代码的空间特征. 按照虚拟地址的顺序,将二进制文件转来的矩阵、操作码与操作数和API向量的矩阵依次填充,并构造一个长度来自虚拟地址数量,宽度等于文件大小/虚拟地址数量,生成三通道矩阵. 最终,生成的图像将具有更丰富的纹理信息,能够更好地反映恶意代码样本与变体之间的相关性.

3.2　坐标注意力机制

由于当前多种反检测技术的使用,以及种种混淆方法使特征的分析和提取越来越复杂,导致恶意代码检测的时效性和准确性大大降低. Squeeze and Excitation Attention Module(SEAM)[19],借助全局池化层计算通道注意力,以较低的计算成本提供准确率提升. 然而,SEAM仅考虑编码恶意代码图像中通道间信息,忽略了位置信息的重要性. Convolutional Block Attention Module(CBAM)[20]在SEAM基础上,引入位置信息,只考虑了特征的局部范围的信息,弱化了通道信息的重要性.

为了更好提取特征信息,本文引入坐标注意力机制,通过位置和空间两种角度提取特征信息,从而加强特征提取能力. 为了获取恶意代码特征之间大范围的

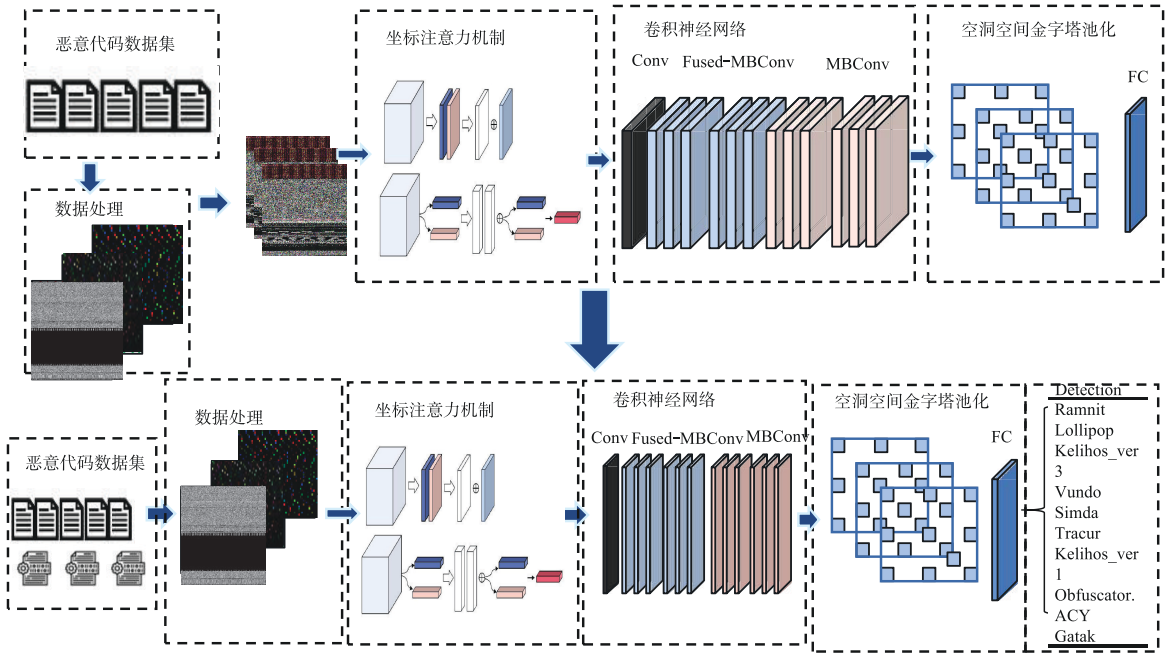


图 1 面向恶意软件变种分类系统架构

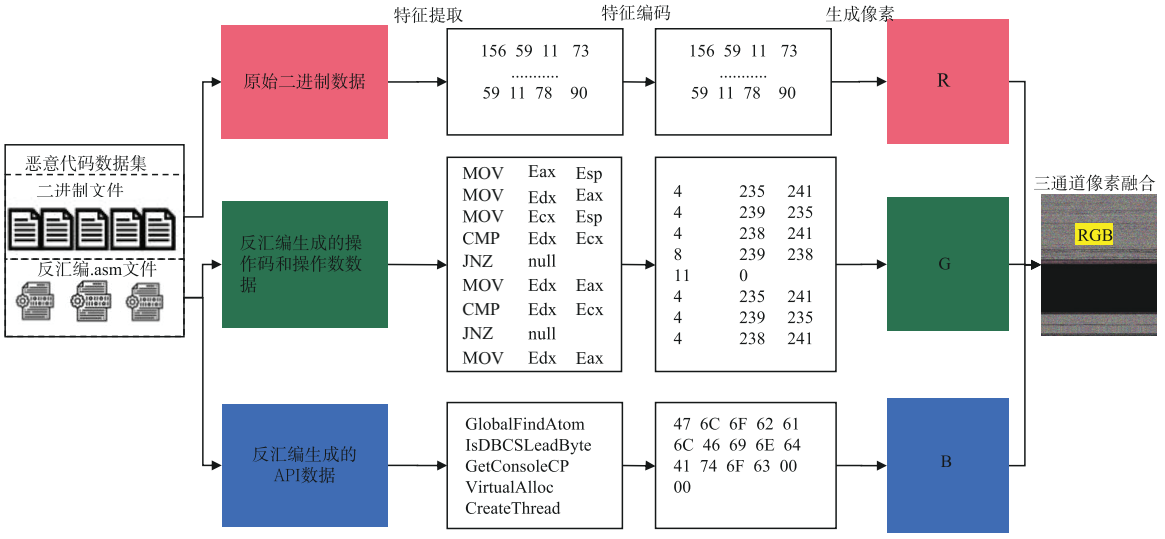


图 2 恶意代码转换为 RGB 图像的流程

空间依赖关系，本文采用 CAM[21] 来具体实现．如图 3 所示．

对输入恶意代码图像特征分别在宽度和高度方向进行一维全局平均池化操作，然后将两个方向的结果拼接在一起，通过 1×1 卷积核映射 Relu 非线性激活层，再分别将两个方向特征分离，并分别通过 1×1 卷积核映射，最终与输入特征做哈达玛积[22]．这样特征图同时具备宽度和高度方向特征的远距离依赖关系能力．

通过加入坐标注意力机制可以获得特征图的空间权重，引导网络获取局部特征及更广泛的空间特征，同时相对 CBAM 减少计算量，进而加速收敛．

3.3 卷积神经网络模块

在传统网络模型中，每个阶段的深度和宽度都是同等放大的．由于每个阶段对网络的训练速度以及参数数量的贡献并不相同，直接使用同等缩放的策略并不合理．本文中参考 Efficientnetv2 模型[23] 采用了非均匀的缩放策略来缩放模型，改进渐进学习方法．该方法会根据训练图像的尺寸动态调节正则方法，用以提升训练速度和准确率，结构如表 3 所示．

3.4 空洞空间金字塔池

池化可以对数据进行降维，从而降低了计算的复杂性，但失去了特征间精确关系信息．然而，传统神经网络要求输入数据固定大小，导致在处理不同结构、大

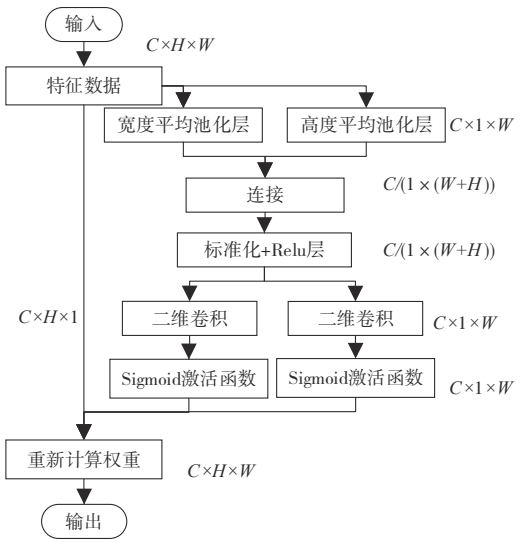


图 3　坐标注意力机制

表3　改进的CNN结构

步骤	操作	步长	卷积层数
0	Coordinate-Attention	1	1
1	Conv3×3	1	1
2	Fused-MBConv1,k3×3	1	2
3	Fused-MBConv4,k3×3	2	4
4	MBConv4,k3×3,SE0.25	2	6
5	MBConv4,k3×3,F0.25	1	9
6	Conv1×1& ASPP& FC	—	1

小时产生局限性.

为了解决上述问题,在模型中引入了空洞空间金字塔池层. 它一种网络结构,接收不同大小的图像,并输出固定维度的向量. ASPP[24,25]由空洞卷积和空间金字塔池(Spatial Pyramid Pooling,SPP)的结合,在多个尺度上捕捉上下文信息,从而实现更准确的分类. 空洞卷积对于输出y和滤波器w上的每个像素i,对输入x是空洞卷积,如等式所示:

$$y[i]=\sum_{k=1}^{k}x[i+r\times k]w[k] \tag{3}$$

其中速率r决定了采样输入恶意代码图像的步距. 输入x与通过在两个连续滤波器值之间插入$R1$个零而产生的滤波器进行卷积. 通过调节速率r,改变滤光器的感受野.

在ASPP结构中的四个平行卷积层的感受野是互补的,就保证了分布在不同范围的信息可以被取样. 利用来自不同范围的信息,分割网络能够处理尺寸在特定范围内变化的对象. 通过使用ASPP,可以将不同大小的训练样本输入到CNN中而不丢失信息. 可以使用传统恶意软件来训练用于恶意软件变体分类的模型,这充分利用了更多的样本来提高检测精度.

3.5　训练阶段

本文中采用了非均匀的缩放策略来缩放模型改进的渐进学习方法. 在训练阶段:首先,将预处理后的RGB图像输入到模型中. 然后,使用损失函数来计算输出与真实值之间的误差. 最后,通过反向传播误差来更新模型的权重. 训练过程中使用Adam优化器和交叉熵损失函数. 模型建立流程如算法1.

算法1　模型算法

输入:数据集

输出:训练模型和分类结果

For $i,j\leftarrow 1$ to k //k is dataset number

$\left\{//M_i=\left\{\mathrm{R}_i,\mathrm{G}_i,\mathrm{B}_i\right\}_{\mathrm{r,g,b}}\right\}$

while $t<$ maxIter do:

for each image $q\in Q$ in batchset:

{

$q_{t+1}(i,j)\leftarrow \mathrm{CA}(\delta(\mathrm{F}(q_{t+1}(i,j))))$

$q_{t+1}(i,j)\leftarrow \mathrm{Fused-MBConv1}(q_{t+1}(i,j))$

$q_{t+1}(i,j)\leftarrow \mathrm{Fused-MBConv4}(q_{t+1}(i,j))$

$q_{t+1}(i,j)\leftarrow \mathrm{MBConv4}(q_{t+1}(i,j))$

$q_{t+1}(i,j)\leftarrow \mathrm{MBConv4}(q_{t+1}(i,j))$

$q_{t+1}(i,j)\leftarrow \mathrm{Conv}(q_{t+1}(i,j))$

$q_t(i,j)\leftarrow q_{t+1}(i,j)$

$q_t(i,j)\leftarrow \mathrm{ASPP}(q_{t+1}(i,j))$

$q_{t+1}(i,j)\leftarrow \mathrm{Dense}(\delta(q_{t+1}(i,j)))$

}

$\mathrm{accuracy}^{t+1}\leftarrow \mathrm{trainModel}(\mathrm{batchset})$

end while

4　数据集和实验设置

4.1　数据集与评价指标

(1) Kaggle

Kaggle[26]恶意软件数据集由一组已知的恶意软件文件组成,代表9个不同的家族,其样本分布如表4.

表4　样本集的数量分布

家族名称	训练样本数	类型
Ramnit	1 541	Worm
Lollipop	2 478	Adware
Kelihos_ver3	2 942	Backdoor
Vundo	475	Trojan
Simda	42	Backdoor
Tracur	751	TrojanDownloader
Kelihos_ver1	398	Backdoor
Obfuscator.ACY	1 228	Any kind of obfuscated malware
Gatak	1 013	Backdoor

（2）DataCon

DataCon[27]所使用数据样本均为现网捕获的真实数据，包含了大量加密、混淆等大量真实恶意代码样本. 其中包括7 896个挖矿样本与15 759个其它家族的恶意样本共23 655个恶意代码样本. 本文使用DataCon数据集作为对比实验，验证本文方法在真实数据集上的有效性.

（3）评价指标

实验评价选用了准确率Accuracy、精确率Precision、召回率Recall和F_1-score等四个指标. 准确率、精确率、召回率和F_1-score的定义：

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{4}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{5}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{6}$$

$$F_1\text{-score}=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{7}$$

4.2 实验对比

考虑到由对抗技术的不断发展，使得变种恶意代码与传统恶意代码的差异增加，从两个方面对方法进行了评估. 一方面，使用Kaggle数据集按照比例为8∶2划分为训练集和测试集，证明了该方法的泛化能力和有效性. 另一方面，使用现网捕获到的恶意代码DataCon数据集，验证了恶意家族变种首次出现时，验证所提出方法在分类恶意软件变种样本上的有效性.

4.2.1 对恶意软件可视化方法的评估

为了证明本文提出的RGB可视化方法的有效性，将其与一些优秀论文中发表的其他可视化方法进行了比较，这些方法的细节见表5.

表5 其它论文中可视化方法的细节

名称	方法概述	相关论文	描述
M1	灰度图	文献[9]	旋转、缩放等归一化方法缩放图,使输入样本尺寸一致
M2	低维灰度共生矩阵	文献[10]	利用低维灰度共生矩阵提取纹理特征和颜色矩阵提取颜色特征,全局与局部特征相结合
M3	Word2Vec多通道方法	文献[15]	基于样本二值灰度图,结合汇编指令级特征和字节级特征Word2Vec的多通道方法

为了证明所提出的可视化方法能够在检测模型上取得令人满意的性能，与恶意软件变种分类方法领域中最常用的三种模型进行了对比测试，如ResNet[28]、DesNet[29]、Xception[30]等，结果见图4.

从图4可以看出，当使用Kaggle数据集时，本文提出的可视化方法分类效果最好. 与M1和M3的灰度图像方法相比，RGB图像包含了更多的信息，能够增强恶意软件特征的差异性，有效避免了训练过程中的过拟合问题. 虽然M2也使用RGB图像来表示恶意软件，但是用不同颜色表示不同特征，这忽略了特征之间的相关性. 与这些方法不同，本文提出的可视化方法从二进制代码、汇编指令和数据和API三个维度提取特征，并通过恶意代码中的虚拟地址建立特征之间的关系. 该方法更好地选择和表现了恶意软件的局部特征，可以针对恶意软件变种实现出色的分类准确率.

当恶意软件变种首次被检测分类时，基于现有的恶意软件分类方法，很难达到较高的准确率. 为了解决这一问题，本文使用现网捕获的DataCon样本来验证模型的效果，实验结果如图5所示.

与图4相比准确率明显下降，主要由于DataCon数据集样本引入更多加密、混淆、变种的恶意软件，导致准确率较低. 实验结果充分证明该方法考虑了恶意软件变种与原体之间的内在关联，从操作码及操作数、API调用等关系多个方面，分析和探索恶意软件变种之间与原家族之间的关系.

4.2.2 坐标注意力机制的评价

基于Kaggle和DataCon两个数据集测试坐标注意机制是否能够提高该模型的检测效果. 在这个实验中，坐标注意力机制对特征的影响结果如图6和图7所示.

根据上图混淆矩阵的预测结果分析：坐标注意力机制可以获取更大范围的空间位置信息，而不只是通道特征信息. 主要原因在于传统SEAM和CBAM，在计算特征过程中因只考虑到位置或者空间特征，忽略空间与位置结合计算特征的重要性. 通过加入坐标注意力机制可以获得特征图的空间权重，引导网络获取局部特征，及更广泛的空间特征，进而加速收敛. 因此，通过引入坐标注意力机制的CNN准确率得到了显著提高.

4.2.3 空洞空间金字塔池的评价

在本节中，将在上述具有坐标注意力机制的CNN模型的基础上，继续评估在该模型中引入空洞空间金字塔池后的改进效果，实验结果如图8和图9所示.

从图8中混淆矩阵可以发现，在引入ASPP之前，诸如家族2和家族9的几个恶意软件家族准确性较低. 传统方法必须通过填充或截断将输入图像固定尺寸，产生冗余信息和信息损失. 通过引入ASPP，可以确保变种恶意软件信息的简洁性和完整性，在家族Lollipop和家族Gatak中的分类准确性得以提升，并且避免尽可能冗余信息和信息损失. 其中，在对Vundo、Simda、Tracur、Kelihos_ver1和Obfuscator. ACY家族表现效果更好

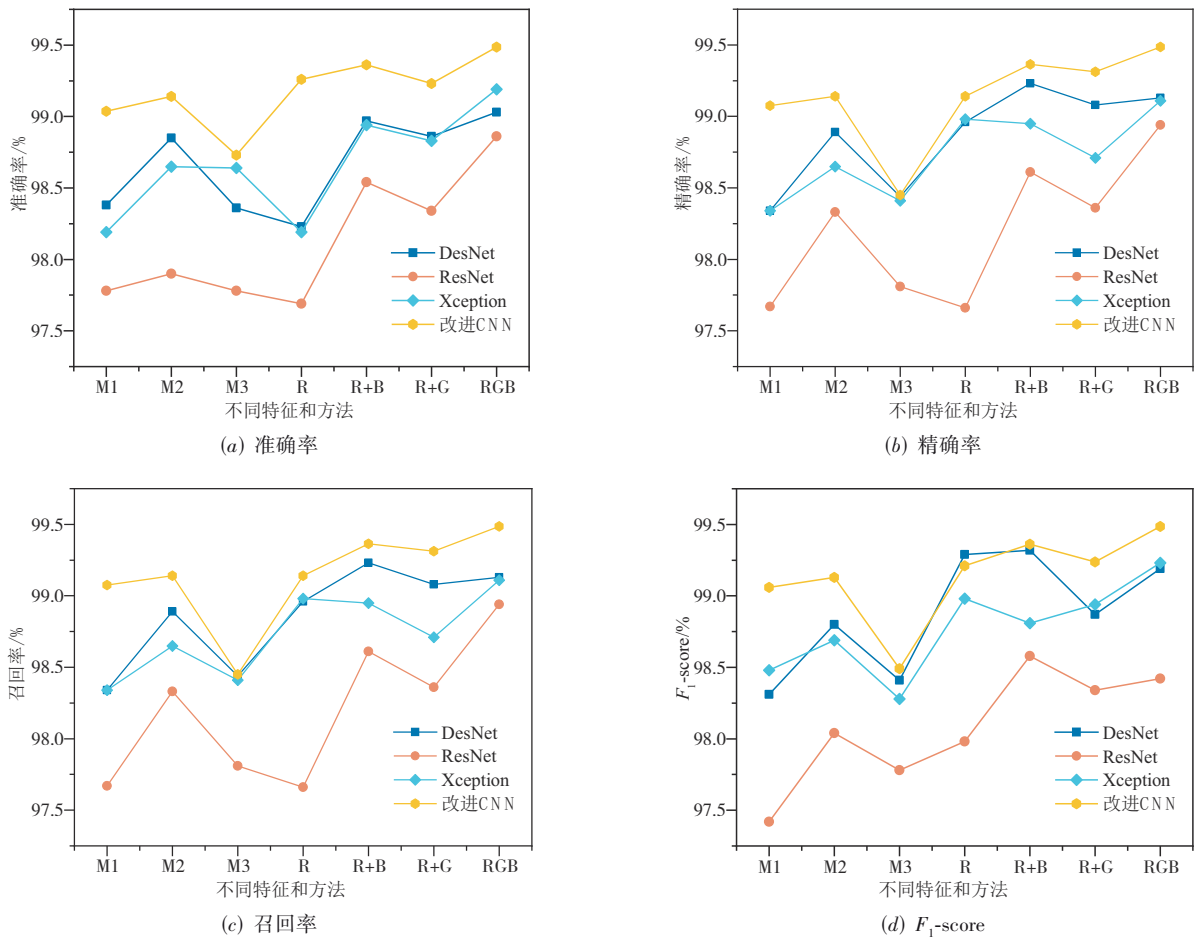


图 4　基于Kaggle数据集的恶意软件可视化方法评估结果

的主要原因在于,RGB特征集成了多种特征混合的优势,保留了变种恶意软件信息的完整性,使得准确率整体表现稳定提升.

4.2.4　空洞空间金字塔池的评价

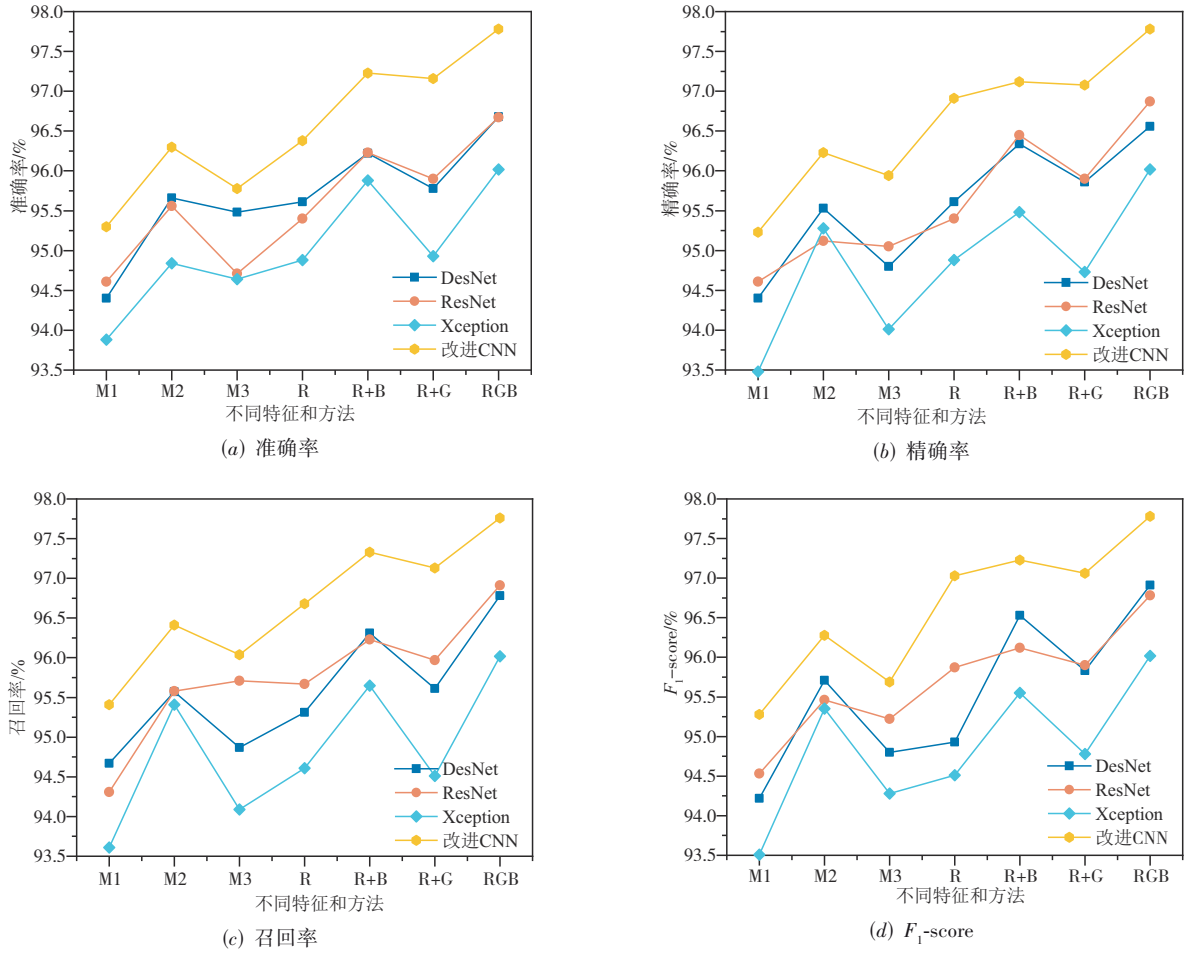
为了验证提出的整体分类系统的有效性,将本文的方法与论文中常用的几种恶意软件变种分类方法进行了比较,结果如表6和表7所示.

表6显示了本文的方法与其它方法在Kaggle数据集上的比较结果. 文献[11]以灰度图像为特征,未考虑二进制代码的语义特征,导致准确率较低,为98.2%. 文献[12]以灰度图像为特征,部分不同家族样本的灰度图像具有较高的相似性,容易产生误判,准确率仅为97.49%. 文献[13]以字节序列和操作码序列为特征,基于签名特征区分恶意代码与变种之间的相似性,但相似性会随着对抗技术的发展而慢慢降低. 文献[15]使用基于马尔可夫图像的VGG16卷积,对特征图像进行归一化使得信息损失和冗余,导致准确率较低. 文献[16]使用基于Word2Vec多通道特征矩阵的LeNet5结构,但是用不同颜色表示不同特征,这忽略了特征之间的相关性.

表7显示了本文的方法在DataCon数据集上与其它方法比较结果,文献[29]以512×512像素大小的灰度

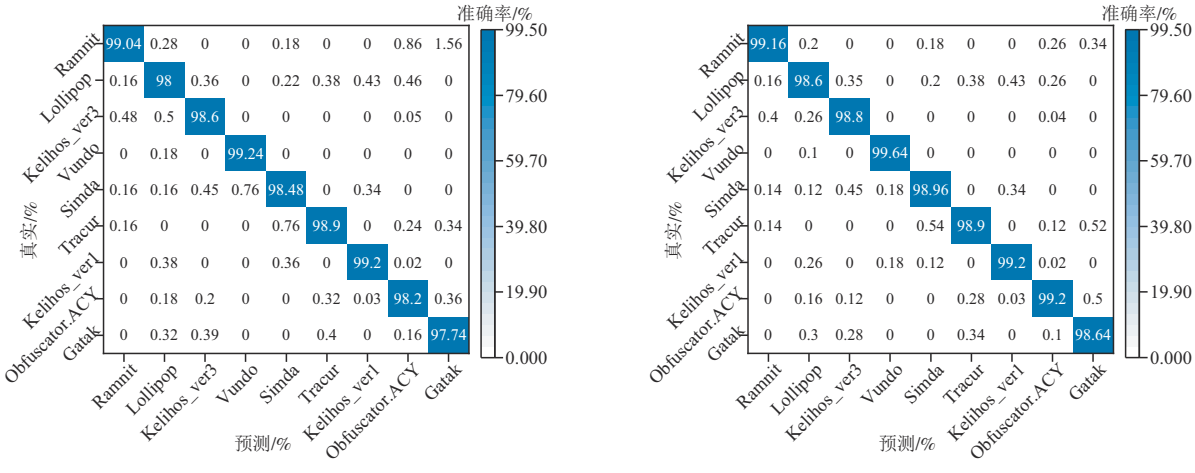
表6　建议的方法与数据集上Kaggle其它方法的比较

方法	方法介绍	Accuracy/%	Precision/%	Recall/%	F_1-score/%
文献[12]	CNN+Gray	97.49	—	—	94.38
文献[11]	CNN+LSTM+Gray	98.20	—	—	95.77
文献[13]	Byte+Opcode	99.24	—	—	98.72
文献[15]	GDMC+Gray	99.26	—	—	—
文献[16]	LeNet5+RGB+Word2Vec	98.76	—	—	—
本文方法	RGB	**99.48**	**99.39**	**99.48**	**99.48**



(a) 准确率 (b) 精确率 (c) 召回率 (d) F_1-score

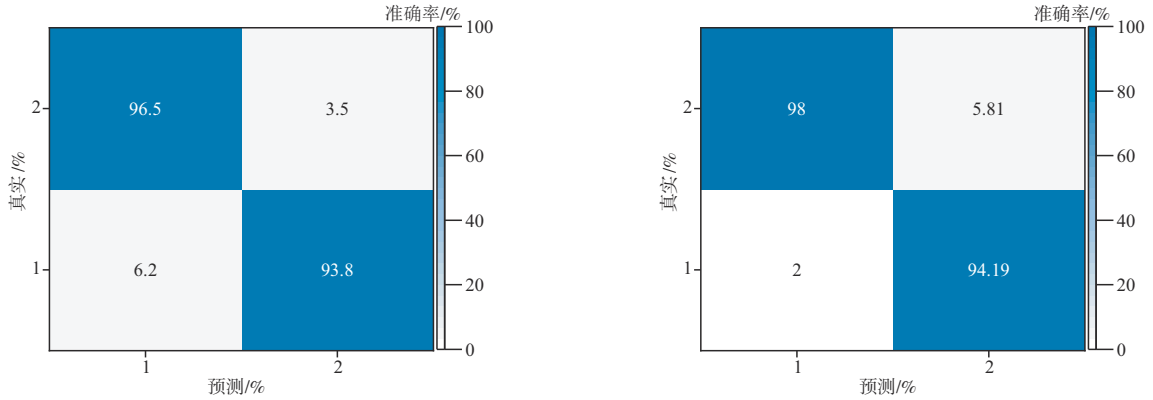
图 5 基于DataCon数据集的恶意软件可视化方法评估结果



(a) CNN (b) CNN+CAM

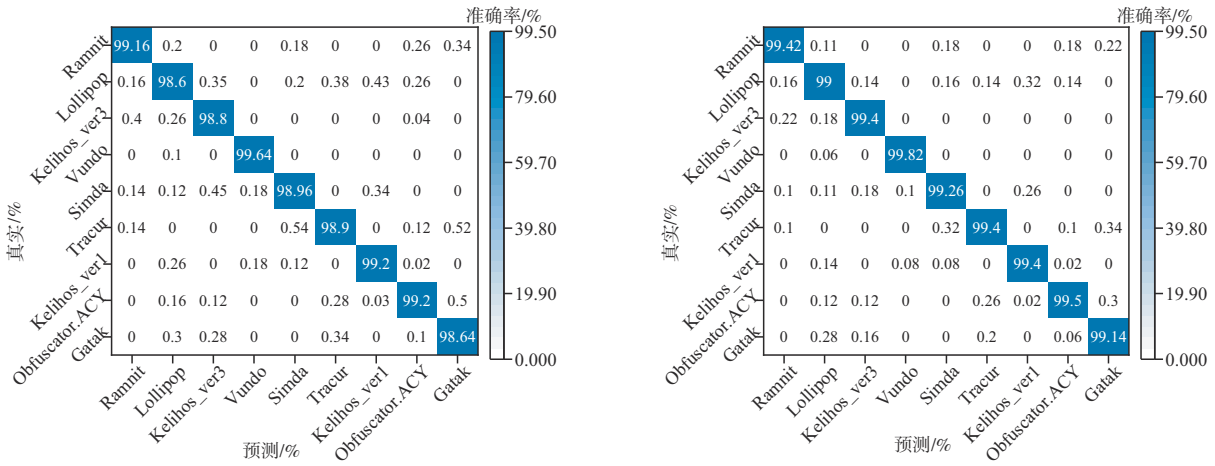
图 6 基于Kaggle数据集的坐标注意力机制影响

图像为特征，未考虑二进制代码的语义特征，导致准确率较低，为96.80%. 文献[31]中包含非PE(Portable Executable)结构、字符串序列、汇编指令、PE结构和CFG(Control Flow Graph)调用关系五种特征，使用由五个组件构成的集成学习分类器，分类器结构复杂，仍有部分样本被误判，准确率为96.99%.



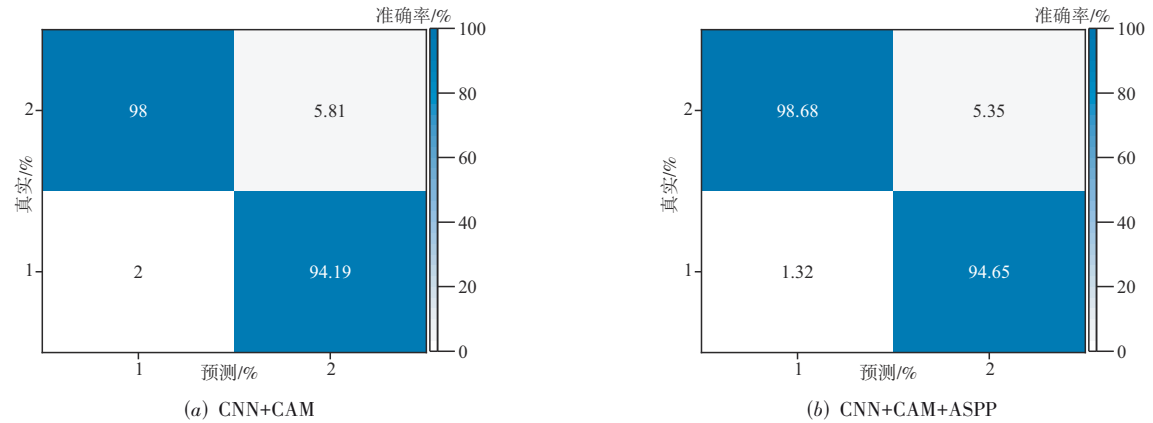
(*a*) CNN　　(*b*) CNN+CAM

图 7　基于DataCon数据集的坐标注意力机制影响



(*a*) CNN+CAM　　(*b*) CNN+CAM+ASPP

图 8　基于Kaggle数据集的空洞空间金字塔池化的影响



(*a*) CNN+CAM　　(*b*) CNN+CAM+ASPP

图 9　基于DataCon数据集的空洞空间金字塔池化的影响

从表6和表7结果可以看出本文的方法比其它分类方法有更高的准确率. 有三个方面原因:

(1) 在应对新型恶意软件时,恶意软件的对抗技术不断发展后,恶意软件变种的文件大小和特征都发生了显著变化,但二进制代码和汇编信息之间的基于虚拟地址的对应关系并没有改变. 与上述方法相比,基

表7 建议的方法与数据集上DataCon其它方法的比较

方法	方法介绍	Accuracy/%	Precision/%	Recall/%	F_1-score
文献[31]	集成学习	96.99	94.05	—	92.19
文献[27]	Gray+CNN	96.80	96.42	96.26	97.38
本文方法	RGB	**97.78**	**97.80**	**97.76**	**97.78**

于这点,本文的可视化方法在多种的特征之间通过虚拟地址建立了联系,因此使用的特征可以更好地表示变化后的特征.

(2) 在模型设计方面,传统分类方法仅考虑了部分特征关系,忽略了全局特征关系. 在特征处理阶段,使用的操作码操作数序列和API调用关系,从特征可视化角度考虑了恶意软件之间的逻辑关系. 而本文使用的坐标注意力方法从模型设计的角度考虑了恶意软件之间的功能和逻辑关系.

(3) 在应对因归一化导致信息冗余或者缺失的挑战方面,ASPP的加入可以将不同大小的恶意图像输入到模型中进行训练和分类,而不是使用零填充或截断,避免了信息的缺失.

5 总结

本文设计实现了一种基于RGB图像获取特征的语义关系,结合ASPP和CA算法对特征增强,提出了基于改进CNN和行为关系特征的恶意代码分类方法. 通过对恶意代码变体的静态分析,避免了归一化导致信息损失,有效提高了恶意代码变体的分类准确率和泛化能力,在恶意代码测试集上验证了本方法的有效性,并为Windows恶意代码分类提供了理论和技术支持.

然而,恶意软件已经成为物联网最严重的安全威胁之一. 针对恶意代码从传统网络向物联网的传播和物联网内部的传播引起的一系列问题,恶意软件变种的分类在物联网领域的应用需要更进一步的探索.

参考文献

[1] MORGAN, Top S. 5 facts cybersecurity, figures and statistics for 2018[R/OL]. [2018-05-05]. https://www.csoonline.com/article/3153707/security/top-5-cybersecurity-facts-figures-and-statistics.html.

[2] Enterprise Symantec. 2018. Internet Security Threat Report 2018[R/OL]. [2019-06-15]. https://www. symantec. com/content/dam/symantec/docs/reports/istr-23-2018-en.pdf.

[3] KHOSHBARFOROUSHHA A, RANJAN R, GAIRE R, et al. Distribution based workload modelling of continuous queries in clouds[J]. IEEE Transactions on Emerging Topics in Computing, 2016, 5(1): 120-133.

[4] TSOCHEV G, TRIFONOV R, NAKOV O, et al. Cyber security: Threats and challenges[C]//2020 International Conference Automatics and Informatics(ICAI). Varna: IEEE, 2020: 1-6.

[5] NATARAJ L, YEGNESWARAN V, PORRAS P, et al. A comparative assessment of malware classification using binary texture analysis and dynamic analysis[C]//Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence. Chicago: ACM , 2011: 21-30.

[6] NATARAJ L, KARTHIKEYAN S, JACOB G, et al. Malware images: visualization and automatic classification[C]//Proceedings of the 8th International Symposium on Visualization for Cyber Security. New York: ACM, 2011: 1-7.

[7] SHAID S Z M, MAAROF M A. Malware behavior image for malware variant identification[C]//2014 International Symposium on Biometrics and Security Technologies(ISBAST). Kuala Lumpur, Malaysia: IEEE, 2014: 238-243.

[8] HAN K S, LIM J H, KANG B, et al. Malware analysis using visualized images and entropy graphs[J]. International Journal of Information Security, 2015, 14(1): 1-14.

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

[10] FU J, XUE J, WANG Y, et al. Malware visualization for fine-grained classification[J]. IEEE Access, 2018, 6: 14510-14523.

[11] LE Q, BOYDELL O, NAMEE B MAC, et al. Deep learning at the shallow end: Malware classification for non-domain experts[J]. Digital Investigation, 2018, 26: S118-S126.

[12] VU D L, NGUYEN T K, NGUYEN T V, et al. A convolutional transformation network for malware classification [C]//2019 6th NAFOSTED Conference on Information and Computer Science (NICS). Hanoi, Vietnam: IEEE, 2019: 234-239.

[13] GIBERT D, MATEU C, PLANES J, et al. Using convolutional neural networks for classification of malware represented as images[J]. Journal of Computer Virology and Hacking Techniques, 2019, 15(1): 15-28.

[14] GIBERT D, MATEU C, PLANES J. Orthrus: A bimodal learning architecture for malware classification[C]//2020 International Joint Conference on Neural Networks (IJCNN). Glasgow, UK: IEEE, 2020: 1-8.

[15] YUAN B, WANG J, LIU D, et al. Byte-level malware classification based on Markov images and deep learning [J]. Computers & Security, 2020, 92: 101740.

[16] QIAN Y, JIANG Q, JIANG Z, et al. A multi-channel vi-

sualization method for malware classification based on deep learning[C]//2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE. Rotorua, New Zealand: IEEE, 2019: 757-762.

[17] LI Q, MI J, LI W, et al. CNN-based malware variants detection method for internet of things[J]. IEEE Internet of Things Journal, 2021, 8(23): 16946-16962.

[18] AMER E, ZELINKA I. A dynamic Windows malware detection and prediction method based on contextual understanding of API call sequence[J]. Computers & Security, 2020, 92: 101760.

[19] FUCHS F, WORRALL D, FISCHER V, et al. SE(3)-transformers: 3D roto-translation equivariant attention networks[J]. Advances in Neural Information Processing Systems, 2020, 33: 1970-198

[20] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer. 2018: 3-19.

[21] CHENG S, WANG L, DU A. Asymmetric coordinate attention spectral-spatial feature fusion network for hyperspectral image classification[J]. Scientific Reports, 2021, 11(1): 1-17.

[22] KIM J H, ON K W, LIM W, et al. Hadamard product for low-rank bilinear pooling[C]//The 5th International Conference on Learning Representations (ICLR). New York: ACM, 2018: 1-7.

[23] TAN M, LE Q. Efficientnetv2: Smaller models and faster training[C]//2019 International Conference on Machine Learning(ICML). Cham: Springer, 2021: 10096-10106.

[24] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[25] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer. 2018: 801-818.

[26] RONEN R, RADU M, FEUERSTEIN C, et al. Microsoft Malware Classification Challenge 2018[EB/OL]. [2019-05-29]. https://doi.org/10.48550/1802/10135.

[27] Qian Xin Technology Research Institute. DataCon: Multi-domain large-scale competition open data for security research[EB/OL]. [2020-08-25]. https://DataCon. qianxin. com/opendata.

[28] 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. Las Vegas: IEEE, 2016: 770-778.

[29] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 4700-4708.

[30] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1251-1258.

[31] 杨望, 高明哲, 蒋婷. 一种基于多特征集成学习的恶意代码静态检测框架[J]. 计算机研究与发展, 2021, 58(05): 1021-1034.
YANG W, GAO M Z, JIANG T. A static detection framework of malware based on multi feature ensemble learning[J]. Journal of Computer Research and Development, 2021, 58(05): 1021-1034. (in Chinese)

作者简介

轩勃娜　女,1991年2月出生于陕西省兴平市. 现为空军工程大学防空反导学院硕士. 主要研究方向为恶意代码分类.
E-mail: afeunbx219318@163.com

李　进　男,1971年9月出生于陕西省西安市. 1988年毕业于空军工程大学电电子学系. 现为空军工程大学副教授,从事地对空导弹指挥控制系统的网络安全.
E-mail: ljlxls@163.com