抗混淆的恶意代码图像纹理特征描述方法

刘亚姝[1,2]，王志海[1]，严寒冰[3]，侯跃然[4]，来煜坤[5]

(1. 北京交通大学计算机与信息技术学院，北京 100044；2. 北京建筑大学电气与信息工程学院，北京 100044；
3. 国家计算机网络应急技术处理协调中心，北京 100029；4. 北京邮电大学网络技术研究院，北京 100876；
5. 卡迪夫大学计算机科学与信息学院，英国 卡迪夫，CF24 3AA）

摘 要：将图像处理技术与机器学习方法相结合是恶意代码可视化研究的一个新方法。在这种研究方法中，恶意代码灰度图像纹理特征的描述对恶意代码分类结果的准确性影响较大。为此，提出新的恶意代码图像纹理特征描述方法。通过将全局特征（GIST）与局部特征（LBP 或 dense SIFT）相融合，构造抗混淆、抗干扰的融合特征，解决了在恶意代码灰度图像相似度较高或差异性较大时全局特征分类准确性急剧降低的问题。实验表明，该方法与传统方法相比具有更好的稳定性和适用性，同时在较易混淆的数据集上，分类准确率也有了明显的提高。

关键词：恶意代码可视化；图像纹理；特征描述符；恶意代码分类

中图分类号：TP393

文献标识码：A

doi: 10.11959/j.issn.1000-436x.2018227

Method of anti-confusion texture feature descriptor for malware images

LIU Yashu[1,2], WANG Zhihai[1], YAN Hanbing[3], HOU Yueran[4], LAI Yukun[5]

1. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China
2. School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture，Beijing 100044, China
3. National Computer Network Emergency Response Technical Team/Coordination Center of China,Beijing 100029, China
4. Institute of Network Technology, Beijing University of Posts and Telecommunication,Beijing 100876,China
5. School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, UK

Abstract: It is a new method that uses image processing and machine learning algorithms to classify malware samples in malware visualization field. The texture feature description method has great influence on the result. To solve this problem, a new method was presented that joints global feature of GIST with local features of LBP or dense SIFT in order to construct combinative descriptors of malware gray-scale images. Using those descriptors, the malware classification performance was greatly improved in contrast to traditional method, especially for those samples have higher similarity in the different families, or those have lower similarity in the same family. A lot of experiments show that new method is much more effective and general than traditional method. On the confusing dataset, the accuracy rate of classification has been greatly improved.

Key words: malware visualization, image texture, feature descriptors, malware classification

1 引言

执行恶意的行为或攻击的软件简称为恶意代码。由于代码自动生成工具的出现和大量攻击代码的公开，恶意代码制作者大量使用可重用模块及自动化变形技术，使得恶意代码数量呈爆发式增长的

收稿日期：2017–10–12；**修回日期：**2018–10–26

通信作者：王志海，zhhwang@bjtu.edu.cn

基金项目：国家自然科学基金资助项目（No.U1736218,No.61672086）；国家重点研发计划基金资助项目（No.2018YFB0803604）

Foundation Items: The National Natural Science Foundation of China (No.U1736218, No.61672086), The National Key Research and Development Program of China (No.2018YFB0803604)

势头，人们普遍认为恶意软件的增长速度远远超过合法应用程序的增速[1]。面对层出不穷的恶意代码威胁和攻击，安全分析人员和研究者已经提出了许多恶意代码的检测技术，但是如何快速、准确地识别、分类恶意代码仍然是这个领域的研究热点。

恶意代码分析技术主要分为静态分析与动态分析两类方法。静态分析是指恶意代码不被实际运行，通过分析恶意代码文件，识别恶意代码的种类和可能造成的危害。动态分析是指在受保护的虚拟环境（如 VMware）中实际运行需要分析的恶意样本，在恶意代码执行过程中分析记录其动态行为特性，针对这些代码表达出来的行为，分析和判断恶意样本的危害级别，为恶意代码样本的识别和清除提供依据。动态分析技术不仅受模拟环境和触发条件等限制，而且随着恶意代码技术的发展，恶意代码的反调试能力不断增强，这大大增加了动态分析的难度。

研究人员已经提出了许多静态分析的方法。其中，基于特征码的分析技术[2-3]被广泛应用到病毒查杀工具中。但是随着技术的改进，出现了具有多态变种能力的恶意代码，能够躲避静态特征码的扫描。因此为了对抗恶意代码的变化，出现了基于行为的分析技术[4]、基于语义分析的方法[5-6]、基于操作码的分析[7-9]等。

在众多研究方法中，恶意代码可视化是一个非常重要的分支。Bonfante[10]提出基于控制流图（CFG, control flow graph）的恶意代码特征表示；Cesare[11]提出一种快速流图分析方法,可以检测加了分组或者多态的样本；Kinable[12]提出了基于调用图（CG, call graph）的方法，能够聚类相似样本，从而快速检测到恶意代码的变体；Trinius[13]将动态分析与树图和线索图相结合来判断样本的恶意性。

随着技术的发展，恶意代码可视化与图像处理技术相结合产生了一个全新的研究视角。但是恶意代码样本产生的图像与普通的图像不同，仅简单地应用现有图像处理的方法，在复杂样本情况下很难得到好的分类结果。

因此，本文提出了一种恶意代码可视化与多特征相融合的分析方法，能够更好地描述恶意代码类别特征，本文主要工作与创新点如下所示。

1) 将恶意代码样本转化为灰度图像，实现了恶意代码的可视化。

2) 提出了将全局特征与局部特征相融合的特征描述新方法，使得新特征更具有抗混淆性。同时，实现了恶意代码的分类问题。

3) 分析了传统方法在复杂数据集分类准确率急剧降低的原因。

4) 通过大量的实验，对比验证本文方法的抗干扰性、抗混淆性和适用性。

2　相关工作

2010 年，Conti[14]提出了将任意二进制文件映射成灰度图像的方法。随后 2011 年 Nataraj 等 [15] 将该思想首次应用于恶意代码的分类中，为恶意代码可视化提出了一种全新的研究方向。

图 1 为某个去掉了 PE 头的十六进制恶意代码“.byte”文件的部分内容。

```
00401000 56 8D 44 24 08 50 8B F1 E8 1C 1B 00 00 C7 06 08
00401010 BB 42 00 8B C6 5E C2 04 00 CC CC CC CC CC CC CC
00401020 C7 01 08 BB 42 00 E9 26 1C 00 00 CC CC CC CC CC
00401030 56 8B F1 C7 06 08 BB 42 00 E8 13 1C 00 00 F6 44
00401040 24 08 01 74 09 56 E8 6C 1E 00 00 83 C4 04 8B C6
00401050 5E C2 04 00 CC CC CC CC CC CC CC CC CC CC CC CC
00401060 8B 44 24 08 8A 08 8B 54 24 04 88 0A C3 CC CC CC
00401070 8B 44 24 04 8D 50 01 8A 08 40 84 C9 75 F9 2B C2
00401080 C3 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
00401090 8B 44 24 10 8B 4C 24 0C 8B 54 24 08 56 8B 74 24
004010A0 08 50 51 52 56 E8 18 1E 00 00 83 C4 10 8B C6 5E
004010B0 C3 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
004010C0 8B 44 24 10 8B 4C 24 0C 8B 54 24 08 56 8B 74 24
004010D0 08 50 51 52 56 E8 65 1E 00 00 83 C4 10 8B C6 5E
004010E0 C3 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
004010F0 33 C0 C2 10 00 CC CC CC CC CC CC CC CC CC CC CC
00401100 B8 08 00 00 00 C2 04 00 CC CC CC CC CC CC CC CC
00401110 B8 03 00 00 00 C3 CC CC CC CC CC CC CC CC CC CC
00401120 B8 08 00 00 00 C3 CC CC CC CC CC CC CC CC CC CC
00401130 8B 44 24 04 A3 AC 49 52 00 B8 FE FF FF FF C2 04
00401140 00 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
00401150 A1 AC 49 52 00 85 C0 74 16 8B 4C 24 08 8B 54 24
00401160 04 51 52 FF D0 C7 05 AC 49 52 00 00 00 00 00 B8
00401170 FB FF FF FF C2 08 00 CC CC CC CC CC CC CC CC CC
00401180 6A 04 68 00 10 00 00 68 68 BE 1C 00 6A 00 FF 15
```

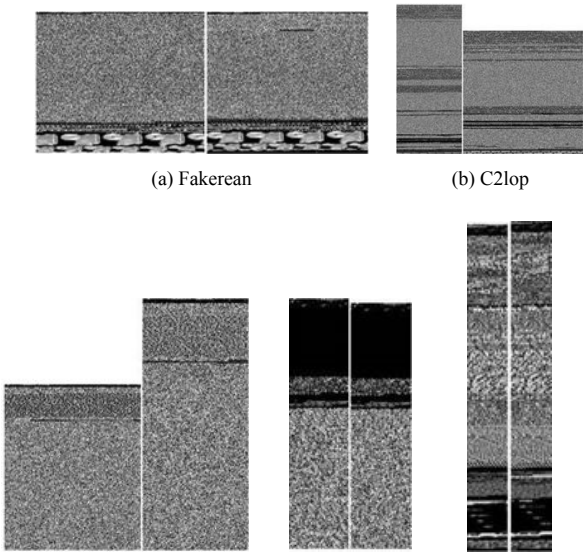
图 1　恶意代码“.byte”文件示例

根据 Nataraj[15]提出的方法，一个恶意代码样本按照每 8 位二进制串对应 1 位十进制数的规则进行转换，得到[0, 255]之间的无符号整数向量。“0”对应黑色、“255”对应白色，因此转换二进制串得到的无符号整数向量能够对应到灰度图像上。但是因为图像是有高和宽的，而无符号整数向量是没有宽度和高度的，因此需要将一维向量转换为二维向量。一般的做法是预先按照样本文件的大小设定图像的宽度，而图像的高度则随着文件大小而变化。本文按照表 1 所示的方式设置图像的宽度，将一个二进制恶意代码可执行文件转化为对应的灰度图像。

表 1 图像宽度的设定标准

文件大小	图像宽度
<10 KB	32
10~ 30 KB	64
30~ 60 KB	128
60~ 100 KB	256
100~ 200 KB	384
200~ 500 KB	512
500 ~ 1000 KB	768
>1 000	1 024

图 2 为按照上述方法得到的 5 个恶意代码家族灰度图像实例。



(a) Fakerean (b) C2lop

(c) Allaple (d) Agent.fyi (e) Lolyda.aa1

图 2 恶意代码家族灰度图像实例

图 2 显示出恶意代码同一家族的图像纹理相似度高、不同家族间的纹理差异大的特点。Nataraj[15]用 GIST 方法提取图像特征、K-近邻方法（KNN, K-nearest neighbor）分类恶意代码图像，取得了非常好的分类效果。

2015 年，Han 等[16]在 Nataraj 方法的基础上通过熵图判断恶意代码的相似性，改进了恶意代码灰度图像纹理特征提取方法以及相似度度量策略。

2018 年 Yan 等[17]基于 LBP 算法提出了改进的恶意代码图像特征提取办法，提高了分类准确性。

不同于上述已有工作，本文提出了一种恶意代码图像特征融合的办法——在全局特征（GIST）的基础上融合局部特征（LBP 或 dense SIFT），构造更健壮的特征描述符，从而解决 Nararaj 方法在某些相似度过高或者差异性较大的家族上分类准确率急剧降低的问题。

3 融合特征的表示

3.1 恶意代码全局特征提取

GIST 方法[18-19]是常用的图像全局特征提取办法之一，它是基于 Gabor 滤波器组进行的。Gabor 滤波器组由多个不同方向和角度的 Gabor 滤波器组成。通过控制尺度和平移因子对 Gabor 函数进行伸缩和平移变换，可以得到一组 Gabor 滤波器组，如式(1)所示。

$$g_{mn}(x,y)=a^{-m}g_{uv}(x',y') \tag{1}$$

其中，$a>1$，a^{-m} 为母小波的膨胀因子，m、n 分别为 Gabor 滤波器组的尺度数和方向数。

GIST 方法首先需要将一幅灰度图像$f(x,y)$划分成 $k\times k$ 的规则网格块，每个网格块按照行依次记作 $p_1,p_2,\cdots,p_t$，其中，t 为网格块的数目。

随后，每一个网格块分别用 m 个尺度和 n 个方向的 Gabor 滤波器进行卷积滤波，则每个网格块经过各通道的滤波后，将卷积结果级联，得到该网格块图像的局部 GIST 特征，如式(2)所示。

$$G_i(x,y)=\underbrace{cat}_{n_c}(f(x,y)*g_{mn}(x,y)) \tag{2}$$

其中，$i=1,\cdots,t$，$(x,y)\in p_i$，n_c 为 $m\times n$ 个通道，g_{mn} 表示滤波函数，cat 被称为级联运算符，*为卷积运算符。

接下来，对 $G_i(x,y)$ 降维，分别计算 $G_i(x,y)$ 均值并将结果按行组合，将该组合称为全局 GIST 特征，如式(3)所示。

$$G=\left\{\overline{G_1},\overline{G_2},\cdots,\overline{G_t}\right\} \tag{3}$$

本文采用 4×4 规则网格划分图像，4 尺度 8 方向的 Gabor 滤波器，可以获得 512 维 GIST 特征。

3.2 恶意代码局部特征提取

局部二值模式（LBP, local binary pattern）是一种用来描述图像局部纹理特征的算子。它首先由 Ojala 等[20]在 1994 年提出。LBP 算子有旋转不变性的特点，但是由于恶意代码图像不涉及到旋转不变的问题，因此本文采用原始的 LBP 算子定义：在 3×3 的窗口内，以窗口中心像素为阈值，将相邻的 8 个像素的灰度值与其进行比较，若周围像素值大于中心像素值，则标记为 1，否则为 0。

因此，3×3 邻域内的 8 个点经比较可产生 8 位二进制数，转换为十进制数即 LBP 码，即得到该窗口中心像素点的 LBP 值。每个区域的特征值计算方法如式(4)所示。

$$LBPc = \sum_{p=0}^{t-1} s(g_p - g_c)2^p \tag{4}$$

其中，g_c 是邻域内中心点 c 的灰度值，g_p 是邻域内第 p 个像素点的灰度值，t 为邻域内像素点个数。$s(x)$ 函数定义如式(5)所示。

$$s(x) = \begin{cases} 1, x \geqslant 0 \\ 0, x < 0 \end{cases} \tag{5}$$

获得每个区域的 LBP 值后通过直方图得到整幅图像的 LBP 特征 $\vec{v}$，如式(6)所示。

$$\vec{v} = hist(\bigcup_{i=1}^{T} LBP_i) \tag{6}$$

其中，$i = 1,2,\cdots,T$，T 是像素点总数，LBP 为像素点 i 的 LBP 特征表示，hist 表示求解直方图。

尺度不变特征变换（SIFT, scale-invariant feature transform）特征描述算子由 David Lowe[21]于 1999 年提出。SIFT 描述算子是关键点邻域高斯图像梯度统计结果的一种表示。通过对关键点周围图像区域分块，计算块内梯度直方图，生成具有独特性的向量，这个向量是该区域图像信息的一种抽象，具有唯一性。dense SIFT 也是 SIFT 方法的一个变化，它提取图像块中每个位置的 SIFT 特征。

本文采用 8×8 固定大小的窗口作为掩模，以 1 为步长在图像上自左向右、从上到下提取图像的 dense SIFT 的特征，可以得到每一个位置的 SIFT 描述符。每个掩模内按照 4×4 的尺度空间、8 个方向获取梯度信息，所以获得图像块每个位置的 dense SIFT 特征为 128 维向量。

3.3　全局特征与局部特征相融合

从恶意代码家族的灰度图像（如图 2 所示）中可以看到，每个家族的全局相似程度很高而差异体现在局部。因此，在提取恶意代码图像的全局特征的前提下，突出局部特征将能够更好地反映恶意代码的家族特征、更具可分性。因此，本文将 GIST 特征分别与 LBP 特征、dense SIFT 特征实现全局与局部特征相融合，如式(7)所示。

$$H = cat(G, R_p(L)) \tag{7}$$

其中，cat 为级联函数；G 为图像的 Gist 特征；$R_p(L)$ 表示随机取得部分局部特征的结果，如 $R_{0.2}(L)$ 表示随机选取 20%的局部特征参与计算；L 为 LBP 特征或 dense SIFT 特征。

由于 LBP 特征是一维向量，可以直接参与计算，但是按照 3.2 节 dense SIFT 方法获取的特征是二维矩阵为

$$\begin{bmatrix} x_{0,0} & \cdots & x_{0,127} \\ \vdots & \ddots & \vdots \\ x_{i,0} & \cdots & x_{i,127} \end{bmatrix}$$

为了获取局部特征参与计算，需要将 dense SIFT 特征离散化，即建立字典。本文方法是随机选取训练集中 t 个行特征向量作为标准词汇，将 dense SIFT 特征矩阵中的行都映射到与选定的 t 个行特征向量中欧氏距离最近的标准词汇上，得到对应的标号，则有

$$\begin{bmatrix} x_{0,0} & \cdots & x_{0,127} \\ \vdots & \ddots & \vdots \\ x_{i,0} & \cdots & x_{i,127} \end{bmatrix} \rightarrow \begin{bmatrix} k_i \\ \vdots \\ k_j \end{bmatrix}$$

其中，k_i，k_j 均为 dense SIFT 特征矩阵中行向量被映射到 t 个行特征向量后得到的标号，$i, j \in (1,\cdots,t)$。计算离散化后的 dense SIFT 统计直方图，得到一维特征向量 $\vec{L}$，如式(8)所示。

$$\vec{L} = hist(\vec{K}) \tag{8}$$

其中，$\vec{K}$ 为图像离散化后得到的行标号的集合。将式(8)的结果用于式(7)，可以获得 GIST 特征与 dense SIFT 特征的融合结果。

3.4　算法设计

根据 3.3 节，本文设计了算法 1 和算法 2 以获得融合特征的分类结果。分类方法采用了 KNN 和随机森林(RF, random forest)。

算法 1　GIST 与 LBP 特征融合分类算法

输入　恶意代码图像数据集 D

输出　恶意代码分类准确率

1) 提取恶意代码图像 GIST 特征 G

2) 提取恶意代码图像 LBP 特征 $\vec{v}$

3) 选取融合参数 p,得到融合特征 H

4) 训练 KNN、RF 分类器，获得分类参数

5) 分类恶意代码，输出分类准确率

算法 2　GIST 与 dense SIFT 特征融合分类算法

输入　恶意代码图像数据集 D

输出　恶意代码分类准确率

1) 提取恶意代码图像 GIST 特征 G

2) 提取恶意代码图像 dense-SIFT 特征

3) dense-SIFT 特征离散化得到 $\vec{L}$

4) 选取融合参数 p,得到融合特征 H

5) 训练 KNN、RF 分类器，获得分类参数

6) 分类恶意代码，输出分类准确率

4 实验与结果分析

按照算法 1 和算法 2，本文在 3 个数据集上完成了实验。数据集分别来自文献[15]中使用的数据集（简称为 NDA，包括 25 个家族 9 458 张恶意代码灰度图像）、文献[15]——Nataraj 个人网站发布的数据集（简称为 NDB，共有 32 个家族，12 278 张恶意代码灰度图）以及 Antiy 实验室提供的数据集（简称为 Antiy，共有 11 个家族，11 000 个恶意代码“.byte”文件）。

文献[15]是最早提出将二进制文件可视化方法应用于恶意代码分类的，将本文方法与文献[15]方法相比也可起到追根溯源的目的；此外，为了验证本文方法的有效性和可适应性，也与其他改进的可视化方法[16-17]以及常见的非可视化方法——基于操作码（OPCode）的恶意代码分析[9]做了对比。

后续实验安排为：首先对比了 GIST 特征（文献[15]采用的方法）在 NDA 和 NDB 上的分类准确性并分析原因，随后给出应用融合特征的分类结果，证明本文方法与文献[15]方法相比更具有抗干扰的能力。本文方法同时应用到 Antiy 数据集上，与文献[9, 15-17]的结果做了对比，证明本文方法更具有一般性。

4.1 在 NDA 与 NDB 上的实验结果

文献[15]在 NDA 数据集上提取 GIST 特征并进行 KNN 分类，可以得到 0.971 8 的正确率。本文按照该文献中的方法，采用 KNN 与 RF 两种分类方法实现了该文献的实验过程。在实验中，每个参数都进行了 10 次实验并取平均值，实验结果如表 2 所示。随后，在 NDB 数据集上，提取 GIST、dense SIFT 以及 LBP 特征，同样采用 KNN 和 RF 两种分类方法，结果如表 3 所示。

表 2 中结果显示在 NDA 数据集上采用 GIST 特征，KNN 分类器获得最高的分类准确率为 0.98(k=2)；RF 分类器平均准确率为 0.988。这与文献[15]中的结果是一致的。

表 2 在 NDA 数据集上 GIST 特征的分类结果（KNN、RF）

分类器	k	准确率
KNN	1	0.976
	2	0.98
	3	0.978
	4	0.977
	5	0.976
	6	0.978
	7	0.97
	8	0.964
	9	0.965
	10	0.96
RF	*estimator*=15	0.988

表 3 中 KNN 分类器的分类准确率随着近邻数目 k 的递增逐渐降低。当 k=1 时采用 GIST 特征的分类准确率最高为 0.910，远远低于表 2 中的结果。同样采用 RF 分类器分类准确率也只有 0.901。

表 3 不同特征描述方法在 NDB 数据集上分类准确性的比较（KNN、RF）

分类器	k	GIST	dense	LBP
KNN	1	0.910	0.936	0.877
	2	0.905	0.926	0.864
	3	0.900	0.925	0.86
	4	0.900	0.909	0.863
	5	0.897	0.913	0.863
	6	0.883	0.912	0.86
	7	0.884	0.914	0.848
	8	0.896	0.914	0.839
	9	0.884	0.912	0.843
	10	0.899	0.906	0.838
RF	*estimator*=15	0.901	0.922	0.901

此外，表 3 也给出 dense SIFT 与 LBP 特征的分类结果。

经仔细对比 NDA 与 NDB 数据集，可以发现 NDA 是 NDB 的子集，并不包括如图 3 和图 4 所示的恶意代码家族。图 3 为 Luder.B 家族灰度图像，该家族样本文件大小差异较大，因此产生的图片宽度也是大小不一，而且图像纹理特征差异也较大。

而图 4 所示 Benign 家族样本图像中带有图片、图标等图案。NDA 中排除了这些易于混淆和干扰信息较多的恶意代码家族，而 NDB 中包含这些样本。因此，有理由假设这些易于混淆的恶意代码家族影响了 NDB 的分类准确性。

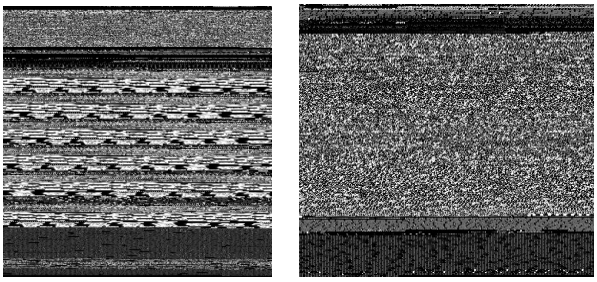


图 3　Luder.B 家族图像

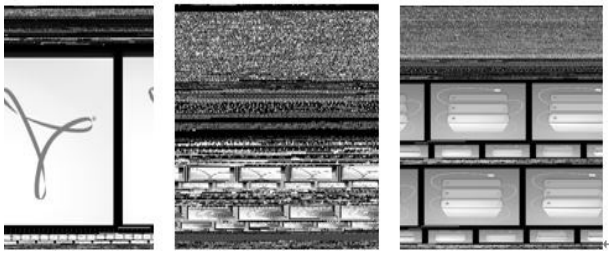


图 4　Benign 家族图像

4.2　假设检验

为了验证 4.1 节的猜测，本文首先从 NDB 中选择了一个包括 9 个家族 2 545 张恶意代码图像的特殊数据集（简称为 NDB-sub 数据集）。NDB-sub 数据集包括家族间特征相似的和家族内特征差异较大的数据，具体信息如表 4 所示。

表 4　　NDB-sub 数据集信息

家族名称	图像数目
Autorun.K	95
Benign	365
Fakerean	381
Luder.B	509
Obfuscator.AD	142
Skintrim.N	80
Virut.A	133
Virut.AC	269
Virut.AK	571

在 NDB-sub 数据集上提取 GIST 特征，采用 RF 分类方法（分类参数为 15），采用十折交叉验证，进行了 10 次实验，结果如表 5 所示。

表 5　GIST 特征在 NDB-sub 数据集上的分类结果

No.	准确率
1	0.914
2	0.886
3	0.910
4	0.867
5	0.906
6	0.882
7	0.906
8	0.894
9	0.906
10	0.89
平均准确率	0.896

对比表 3 与表 5 可以看到，采用 GIST 方法在 NDB 与 NDB-sub 数据集上的分类准确率基本上是相符的，说明表 4 所示的 NDB-sub 数据集中的数据是影响 NDB 数据集分类准确性的主要家族。

表 6 为分类准确率为 0.914 时的混淆矩阵。从此时的混淆矩阵可以看到，测试数据主要在 Benign 与 Luder.B 家族中发生较严重的分类错误问题，这两个家族相互之间以及与 Virut.AK 家族间会发生分类错误的现象。此外， Virtut.A 家族错分到 Virut.AC 类的数据错误率也达到了 0.118，Fakerean 家族也有 0.026 的错误率。

接下来，按照算法 1 进行了测试。在 GIST 特征的基础上增加了 70%的 LBP 特征得到融合特征。当分类参数设置为 25 时，在 NDB-sub 数据集上 RF 的分类结果最好可以达到 0.953，此时的混淆矩阵如表 7 所示。

对比表 6 与表 7 的混淆矩阵可以看到，Fakerean、Virtut.A 家族已经全部分类正确；Benign 与 Luder.B 家族的分类正确率也有了提高。这说明本文提出的恶意代码图像融合特征的方法更具抗混淆和抗干扰能力。

4.3　本文方法在 NDB 数据集上的实验

在 4.2 节中已经说明了本文方法的有效性，为了进一步验证其稳定性和有效性，本文在数据集 NDB 上设计了以下实验。

表 6 GIST 特征在特殊数据集分类中的混淆矩阵

	Autorun.K	Benign	Fakerean	Luder.B	Obfusca-	Skintrim.N	Virut.A	Virut.AC	Virut.AK
Autorun.K	1	0	0	0	0	0	0	0	0
Benign	0	0.667	0	0.303	0	0	0	0	0.03
Fakerean	0	0	0.974	0	0	0	0	0	0.026
Luder.B	0	0.102	0	0.837	0	0	0	0	0.061
Obfusca-	0	0	0	0	1	0	0	0	0
Skintrim.N	0	0	0	0	0	1	0	0	0
Virut.A	0	0	0	0	0	0	0.882	0	0.118
Virut.AC	0	0	0	0	0	0	0	1	0
Virut.AK	0	0	0	0	0	0	0	0	1

表 7 抗混淆新特征在特殊数据集分类中的混淆矩阵

	Autorun.K	Benign	Fakerean	Luder.B	Obfusca-	Skintrim.N	Virut.A	Virut.AC	Virut.AK
Autorun.K	1	0	0	0	0	0	0	0	0
Benign	0	0.75	0	0.208	0	0	0	0	0.042
Fakerean	0	0	1	0	0	0	0	0	0
Luder.B	0	0.082	0	0.898	0	0	0	0	0.02
Obfuscator.AD	0	0	0	0	1	0	0	0	0
Skintrim.N	0	0	0	0	0	1	0	0	0
Virut.A	0	0	0	0	0	0	1	0	0
Virut.AC	0	0	0	0	0	0	0	1	0
Virut.AK	0	0	0	0	0	0	0	0	1

实验 1 按照算法 1，构造 GIST 特征与随机取得 10%、30%、50%、70%、100%的 LBP 特征作为融合特征，采用 RF 分类方法、十折交叉验证分别进行了实验，结果如表 8 所示。

实验中 RF 的分类参数分别选取了 10、15、20、25，每个参数都进行了 10 次实验，取平均准确率。对比本文方法与仅采用Gist、LBP特征的分类准确率，可以看到增加了 LBP 特征后，分类准确率有明显的提高。例如，增加 100%LBP 特征，20 棵树时平均准确率为 0.964，而 GIST 特征的分类准确率只有 0.899。

实验 2 按照算法 2，构造 GIST 特征与随机取得 10%、30%、50%、70%、100%的 dense SIFT 特征作为融合特征，RF 分类结果如表 9 所示。

从表 9 可以看到 GIST 融合了 dense SIFT 特征后分类准确率得到了明显地提高。而且也可以看到，dense SIFT 特征不同的选取比例对分类结果的影响较小。

图 5 为 Gist 特征融合 70% dense SIFT 特征与仅采用 GIST 与 dense SIFT 的分类结果曲线图。这里采用 RF 分类器，参数为 10、15、20 及 25，分别进行了 10 次实验。

图 5 中可以清楚地看到GIST 特征融合了 dense SIFT 特征后每一次的分类结果都是三者中最好的。

从以上实验结果可以看到，本文提出的恶意代码图像特征描述方法在大规模的数据集上也具有较高的分类准确性和稳定性。

4.4 对比与分析

为了进一步验证本文方法的稳定性和适应性，将文献[9, 15-17]以及本文方法分别用于 Antiy 数据集，并对比分类结果。

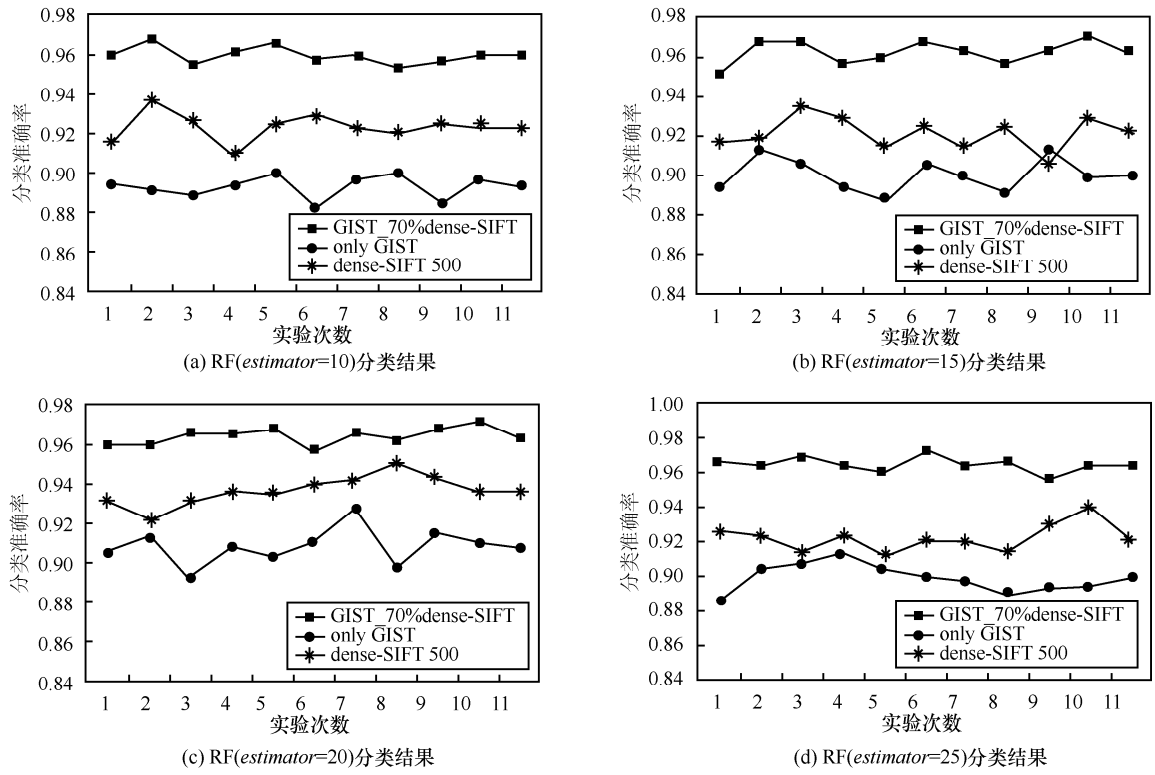
实验 3 恶意代码可视化方法在 Antiy 数据集上的对比。如前所述，文献[15]将恶意代码二进制文件转换为位图后提取 GIST 特征。文献[16]是在文献[15]的基础上提出的恶意代码图像特征表示

表 8　　Gist 特征与 LBP 特征相融合的实验结果

RF	Gist 特征与 LBP 特征相融合					Gist	LBP
	Gist_10%LBP	Gist_30%LBP	Gist_50%LBP	Gist_70%LBP	Gist_100%LBP		
estimator=10	0.954	0.958	0.959	0.957	0.958	0.893	0.899
estimator=15	0.953	0.958	0.961	0.960	0.961	0.901	0.901
estimator=20	0.955	0.960	0.962	0.961	0.964	0.899	0.903
estimator=25	0.957	0.963	0.965	0.962	0.964	0.900	0.904

表 9　　Gist 特征与 dense SIFT 特征相融合的实验结果

RF	Gist 特征与 dense SIFT 特征相融合					Gist	Dense SIFT
	10% dense SIFT	30% dense SIFT	50% dense SIFT	70% dense SIFT	100% dense SIFT		
estimator=10	0.957	0.960	0.960	0.960	0.960	0.893	0.924
estimator=15	0.961	0.962	0.963	0.964	0.964	0.901	0.922
estimator=20	0.963	0.964	0.963	0.964	0.963	0.899	0.931
estimator=25	0.965	0.965	0.965	0.966	0.966	0.900	0.923



(a) RF(*estimator*=10)分类结果　　(b) RF(*estimator*=15)分类结果

(c) RF(*estimator*=20)分类结果　　(d) RF(*estimator*=25)分类结果

图 5　GIST 联合 dense SIFT 特征分类结果

的新方法。在该文献中将二进制恶意代码文件转换为位图后，并没有采用常规的图像特征描述方法，而是计算位图图像每行的熵值，并将熵值表示为熵图。将熵图作为判断二进制恶意代码文件相似性的特征。文献[16]中所表述方法与文献[15]具有相似的准确率，但是与文献[15]相比具有更少的计算量、更快的判别速度。文献[17]改进了文献[15]对恶意代码图像的特征提取方法——采用改进的

LBP 方法（PRICoLBP）以提高特征的辨识性。文献[16]和文献[17]都是针对恶意代码图像特征表示方法的改进，这与本文方法具有相同的研究方向，因此，更具有可比性。

为了对比本文方法与上述文献方法的适用性，本文在 Antiy 数据集进行了实验。实验中，本文方法采用的是 GIST 与 100%LBP 特征相融合以表示恶意代码图像特征。表 10 中文献[15-17]以及本文方法的准确率均是采用 RF 分类方法，分类参数为 15 时的实验结果。文献[16]的准确率是按照该文献提供的熵图相似性比较公式得到的结果，阈值设为 0.75。从表 10 中可以看到，本文方法相对于其他恶意代码可视化方法而言具有更好的分类效果。

表 10 不同文献方法在 Antiy 数据集上分类准确率的比较

特征表示方法	准确率
文献[15]	0.901 7
文献[16]	0.912
文献[17]	0.932 9
本文方法	0.949 8

实验 4 本文方法与恶意代码非可视化方法的比较。如前所述，从恶意代码分类方法上看，针对恶意代码反汇编文件的分析也是静态分析技术的重要研究方向之一。文献[9]是在常用的恶意代码 OPCode 操作码相似性比较的基础上做的改进，取得了很好的恶意代码同源性比较结果。为此，也将本文方法与这种非可视化方法进行了对比。文献[9]需要分析恶意代码的反汇编文件，提取 3-gram 的 Opcode 操作码，获得 simHash 值并配合函数跳转图能够快速判断恶意代码的相似性，并在该文献提供的数据集上溯源准确率可以达到 0.959 9。但是将该方法用于 Antiy 数据集上结果并不理想。

在实验中，需要将 Antiy 数据集中的恶意代码“.byte”文件反汇编，得到对应的 11 000 个“.asm”文件。首先在 20 个样本的实验中（随机选取家族 9 中 10 个样本，另外 10 个样本从其他家族中随机抽取），判别得出 7 个样本属于家族 9，但这 7 个样本中仅有 3 个是真的属于家族 9，误报误判率都很高。随后在 Antiy 数据集全部 11 000 个文件上的实验中发现分类准确率仅有 0.573（这说明在 Antiy 数据集上文献[9]的方法几乎是不可分的）。从表 10 中可以看到本文方法在 Antiy 数据集上也获得了 0.949 8 的分类准确率。因此，本文方法与文献[9]相比更具有适应性。

经过上述实验可以得出以下结论：本文提出的恶意代码图像的全局特征融合局部特征的方法是可行的，能够产生更抗混淆性和抗干扰性的特征向量，对数据集具有更好的适应性和健壮性。

5 结束语

本文主要研究了恶意代码可视化图像的特征描述方法，对比文献[15]中提到的方法，分析其存在的问题，提出了全局特征与局部特征相融合的特征表示方法。此外，本文也与其他方法进行了对比，实验结果表明，在更一般性的数据集上，本文的方法具有更好的适应性、抗干扰性和抗混合性，可以得到更好的分类结果。

参考文献：

[1] 杜敬凯. 二进制恶意代码的同源性分析[D]. 北京: 北京航空航天大学. 2016.
DU J K. Homology analysis of binary malicious code[D]. Beijing: Beihang University. 2016.

[2] SATHYANARAYAN V S, KOHLI P, BRUHADESHWAR B. Signature generation and detection of malware families[C]//Proceedings of Australasian Conference on Information Security and Privacy. 2008:336-349.

[3] ABBAS M F B, SRIKANTHAN T. Low-complexity signature-based malware detection for IoT devices[C]//Proceedings of Applications and Techniques in Information Security. 2017:181-189.

[4] FIRDAUSI I, LIM C, ERWIN A, et al. Analysis of machine learning techniques used in behavior-based malware detection[C]//IEEE International Conference on Advances in Computing. 2010: 201-203.

[5] 王蕊,冯登国,杨轶,等.基于语义的恶意代码行为特征提取及检测方法[J].软件学报,2012, 23(2):378-393.
WANG R, FENG D G, YANG Y, et al. Semantics-based malware behavior signature extraction and detection method[J]. Journal of Software, 2012, 23(2): 378-393.

[6] 任李,潘晓中.基于对象语义的恶意代码检测方法[J].计算机应用研究,2013,30(10):3106-3113.
REN L, PAN X Z. Object-semantics based malware detection method[J]. Application Research of Computers. 2013, 30(10): 3106-3113.

[7] SANTOS I, BREZO F, NIEVES J, et al. Idea: opcode-sequence based malware detection[C] //International Conference on Engineering Secure Software and Systems. 2010: 35-43.

[8] O'KANE P, SEZERAND S, MCLANGHLIN K. Detecting obfuscated malware using reduced opcode set and optimized runtime trace[J]. Security Informatics, 2016, 5(1):2-13.

[9] QIAO Y C, YUN X C, ZHANG Y Z, et al. Fast reused function retrieval method based on simHash and inverted index[C]//The 15th

IEEE International Conference on Trust, Security and Privacy in Computing and Communications.2017: 937-944.

[10] BONFANTE G, KACZMAREK M, MARION JY. Architecture of a morphological malware detector[J]. Computer Virology. 2009, 5(3): 263-270.

[11] CESARE S, XIANG Y. A fast flow graph based classification system for packed and polymorphic malware on the end host[C]//Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications. 2010: 721-728.

[12] KINABLE J, KOSTAKIS O. Malware classification based on call graph clustering[J]. Computer Virology. 2011,7(4): 233-245.

[13] TRINIUS P, HOLS T, GOBEL J, et al. Visual analysis of malware behavior using treemaps and thread graphs[C]//the 6th International Workshop on Visualization for Cyber Security. 2010: 33-38.

[14] CONTI G, BRATUS S, SHUBING A, et al. Automated mapping of large binary objects using primitive fragment type classification[J]. Digital Investigation: The International Journal of Digital Forensics and Incident Response, 2010, 7: S3–S12.

[15] NATARAJ L, KARTHIKEYAN S, JACOB G, et al. Malware images: visualization and automatic classification[C]//The 8th International Symposium on Visualization for Cyber Security. 2011: 21-29.

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

[17] YAN H B, ZHOU H, ZHANG H G. Automatic malware classification via PRICoLBP [J]. Chinese Journal of Electronics, 2018, 27(4): 852-859.

[18] OLIVA A, TORRALBA A. Modeling the shape of the scene: a holistic representation of the spatial envelope[J]. International Journal of Computer Vision. 2001,42(3):145-175.

[19] TORRALBA A, MURPHY A, FREEMAN K P, et al. Context-based vision systems for place and object recognition[C]//International conference on Computer Vision.2003: 273.

[20] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 24(7):971-987.

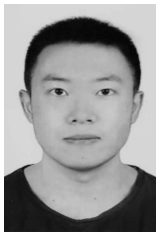
[21] LOWE D G. Object recognition from local scale-invariant features[C]//International Conference on Computer Vision. 1999: 1150-1157.

[作者简介]

刘亚姝（1977−），女，吉林大安人，北京交通大学博士生，主要研究方向为信息安全、数据挖掘。

王志海（1963–），男，河南安阳人，博士，北京交通大学教授、博士生导师，主要研究方向为数据挖掘、机器学习、计算智能。

严寒冰（1975–），男，江西进贤人，博士，国家计算机网络应急技术处理协调中心教授级高工、博士生导师，主要研究方向为信息安全。



侯跃然（1994–），男，内蒙古呼和浩特人，北京邮电大学硕士生，主要研究方向为信息安全、机器学习。

来煜坤（1978–），男，浙江萧山人，博士，英国卡迪夫大学副教授，主要研究方向为计算机视觉、图像处理。