计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210200166

基于 Attention-DenseNet-BC 的恶意软件家族分类方法

李一萌　李成海　宋亚飞　王　坚
空军工程大学防空反导学院　西安 710051
(liyimeng0378@163.com)

摘　要　恶意软件是互联网最严重的威胁之一。现存的恶意软件数据庞大,特征多样。卷积神经网络具有自主学习的特点,可以用来解决恶意软件特征提取复杂、特征选择困难的问题。但卷积神经网络连续增加网络层数会引起梯度消失,导致网络性能退化、分类准确率较低。针对此问题,提出了一种适用于恶意软件图像检测的 Attention-DenseNet-BC 模型。首先结合 DenseNet-BC 网络和注意力机制(attention mechanism)构建了 Attention-DenseNet-BC 模型,然后将恶意软件图像作为模型的输入,通过对模型进行训练和测试得到检测结果。实验结果表明,相比其他深度学习模型,Attention-DenseNet-BC 模型可以取得更好的分类结果。在 Malimg 公开数据集上该模型取得了较高的分类精确率。

关键词:恶意软件;DenseNet-BC 网络;注意力机制
中图法分类号　TP393.08

Method of Malware Family Classification Based on Attention-DenseNet-BC Model Mechanism

LI Yi-meng,LI Cheng-hai,SONG Ya-fei and WANG Jian
Air and Missile Defense College,Air Force Engineering University,Xi'an 710051,China

Abstract　Malware is one of the most serious threats to the Internet. The existing malware has huge data size and various features. Convolutional Neural Network has the features of autonomous learning,which can be used to solve the problems that the feature extraction of malware is complex and the feature selection is difficult. However,in convolutional neural network,continuously increasing the network layers will cause a disappear of the gradient,leading to a degradation of network performance and low accuracy. To solve this problem,an Attention-DenseNet-BC model that is suitable for malware image detection is proposed. First,the Attention-DenseNet-BC model is constructed by combining the DenseNet-BC network and the attention mechanism. Then,the malware images are used as the input of the model,and the detection results are obtained by training and testing the model. The experimental results indicate that compared with other deep learning models,the Attention-DenseNet-BC model can achieve better classification results. A high classification accuracy can be attained with the model based on the malimg public dataset.

Keywords　Malware,DenseNet-BC network,Attention mechanism

1　引言

近年来,随着互联网技术的高速发展,互联网内容服务更加丰富。根据 CNNIC 于 2020 年发布的《中国互联网络发展状况统计报告》,截至 2020 年 6 月,我国网民数量已达 9.40 亿,普及率达到 67.0%[1]。然而,随着互联网用户的增长,随之而来的安全问题也更加严重。根据《2020 年上半年我国互联网网络安全监测数据分析报告》,2020 年上半年,我国捕获计算机恶意程序样本数量约 1815 万个,日均传播次数达 483 万余次,涉及计算机恶意程序家族约 1.1 万余个。我国境内感染计算机恶意程序的主机数量约 304 万台,同比增长 25.7%[2]。恶意软件造成的信息泄露等问题,给人身安全带来巨大隐患,对经济造成巨大损失,故对恶意软件的检测和研究越来越重要。

恶意软件指在系统设备上执行恶意任务的含有病毒、蠕虫和特洛伊木马程序的软件,在未经用户许可的情况下控制用户设备,盗取用户个人信息,破坏了系统的保密性、完整性、可用性。恶意软件检测的基本任务是通过学习到的恶意软件多维度的特征,识别已经出现的恶意软件对网络、系统和用户产生的危害以及可能发生的恶意行为对网络、系统和用户造成的潜在威胁。恶意软件检测的模型一般是由特征提取和训练检测两部分组成。恶意软件检测按检测方式分为动态检测

到稿日期:2021-02-25　返修日期:2021-07-01
基金项目:国家自然科学基金(61703426);陕西省高校科协青年人才托举计划(2019038);陕西省创新能力支撑计划(2019-065)
This work was supported by the National Natural Science Foundation of China(61703426),Young Talents Promotion Program of Shaanxi University Science and Technology Association(2019038) and Innovation Capability Support Plan of Shaanxi Province(2019-065).
通信作者:李成海(lichenghai_ns@163.com)

和静态检测。动态检测指在运行程序样本过程中检测样本，根据样本在运行状态下是否采取了恶意行为来判断该样本是良性软件还是恶意软件，需要消耗的资源较多。静态检测则不需要运行样本，通过反编译、图像化等手段提取样本特征进行检测，所消耗的资源较少。

随着神经网络[3]在各个领域的发展，利用神经网络可以自动学习输入数据特征的优点，研究者们将神经网络引入到网络安全的各个领域以解决恶意软件特征提取复杂、特征选择困难的问题。恶意软件检测研究通过使用深度学习的方法取得了一定的进展，很多静态检测方法使用了神经网络模型作为恶意软件检测或分类的模型。

近年来，国内外学者对该领域的研究大致如下：文献[4]利用自编码网络对恶意软件进行检测，这种方法能够比传统机器学习方法得到更好的检测效果，但是存在无法检测混淆恶意代码的问题。文献[5]对恶意代码进行可视化，将恶意代码家族可执行文件转化为灰度图后，利用 CNN-BiLSTM 网络模型对图像数据集进行检测分类。文献[6]对恶意软件的 3 个特征进行可视化规整，针对其特征图构建了多路卷积神经网络进行恶意软件家族分类。文献[7]在特征图的构建模块中使用了 Bi-GRU 和 Bi-RNN，最后使用三层卷积神经网络对恶意软件家族进行分类。在进行恶意软件家族分类时，部分类别恶意软件图像的相似度高，在网络中学习效果较差，会造成其分类准确率较低。

随着卷积神经网络的发展，不同的网络模型被提出，如 AlexNet[8]，LeNet[9]，VGG[10]，Inception[11-12]，ResNet[13]，FCN[14]，以及递归神经网络 LSTM 和 GRU[15-16]等。为了提高网络模型的分类精度，增加网络模型的深度成为了模型优化的方向之一，然而，模型深度的增加会引起梯度消失的问题，网络性能反而退化，造成分类准确率较低。

为了解决以上卷积神经网络随着网络层数增加引起梯度消失，产生网络性能退化、准确率较低的问题，本文提出了一种结合注意力机制和 DenseNet-BC 网络模型的恶意软件分类方法。DenseNet-BC 网络[17]由于鼓励特征重用，加强了特征传播，可以更好地挖掘图像中的深层特征，而增加注意力机制可以使网络更关注特征层中的有效特征通道，帮助网络获取兴趣区域，减少对非重要信息的关注度。在 DenseNet-BC 网络的基础上进一步提升了网络整体性能，有效缓解了梯度消失的问题，提高了分类准确率。

2 算法模型

2.1 DenseNet-BC 网络

DenseNet 是 Huang 等[17]提出的一种稠密卷积网络，它侧重于从特征重用的角度提升网络性能，在一定程度上缓解了随着深度增加出现的梯度消失问题，其网络结构如图 1 所示。

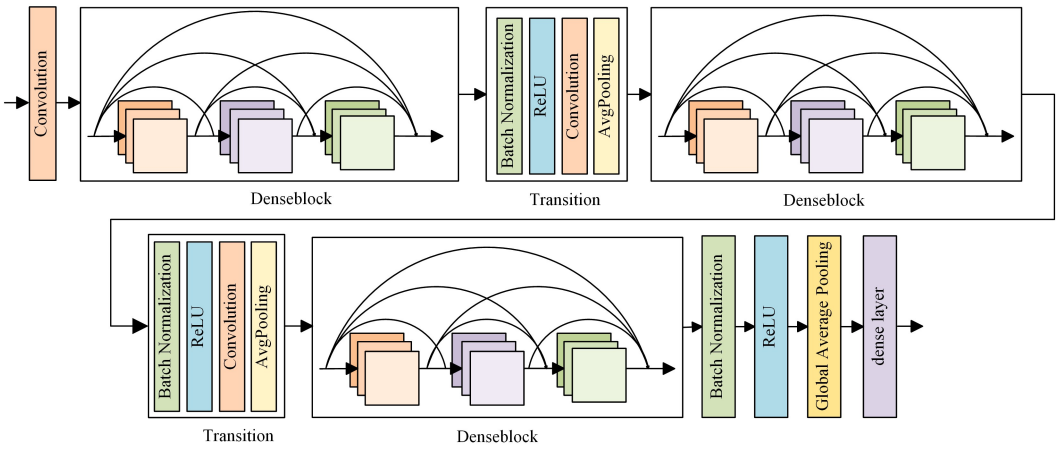


图 1 DenseNet 模型结构图

Fig. 1 DenseNet model structure diagram

2.1.1 网络结构

(1)Denseblock

假设通过卷积网络传递的单个图像 X_0，卷积神经网络由 l 层组成，每一层中实现了一个非线性变换 $H_l(\cdot)$，该非线性变换可以是一系列操作的复合函数，例如 ReLU 函数、卷积操作、池化操作等。

在传统的前馈网络中，将 $l-1$ 层的图像特征传递给 l 层的传递方式为：

$$X_l = H_l(X_{l-1}) \tag{1}$$

ResNet 网络的核心是增加了一个跳跃连接，通过建立前面层与后面层的“短路连接”来减缓梯度消失。它的图像特征传递方式为：

$$X_l = H_l(X_{l-1}) + X_{l-1} \tag{2}$$

在 DenseNet 网络的 Denseblock 结构中，任何两层之间都是直接连接的，即网络的每一层输入都是前面所有层输出的并集(Concatnate 操作)，每一层所学习的特征图也会作为输入被直接传递给其后的所有层。它的图像特征传递方式为：

$$X_l = H_l([X_0, X_1, \cdots, X_{l-1}]) \tag{3}$$

其中，$[X_0, X_1, \cdots, X_{l-1}]$表示分别在第 0 层到第 l 层生成的特征图像。

在 Denseblock 中，各个层的特征图大小一致，可以在通道维度上连接。Denseblock 中的非线性组合函数 $H_l(\cdot)$采用的是 BN+ReLU+3×3 Conv 结构。

假定输入层的特征图的通道数为 k_0，那么 l 层输入的通道数为 $k_0+(l-1)k$，其中 k 为网络的增长率(growth rate)，其实际含义是这层新提取出的特征。

(2)Transition 层

Transition 层连接两个相邻的 Denseblock，它包括一个 1×1 的卷积和 2×2 AvgPooling，其结构为 BN＋ReLU＋1×1 Conv＋2×2 AvgPooling。

2.1.2 减少网络参数的工作

(1)瓶颈层(bottleneck layers)

由于 DenseNet 注重特征重用，越到后面层的输入会越大，因此 Denseblock 内部采用 bottleneck 来减少计算量，并在原有的结构中增加 1×1 Conv，即 BN＋ReLU＋1×1 Conv＋BN＋ReLU＋3×3 Conv，该结构称为 DenseNet-B 结构。加入瓶颈层前后其结构变化如图 2 所示。

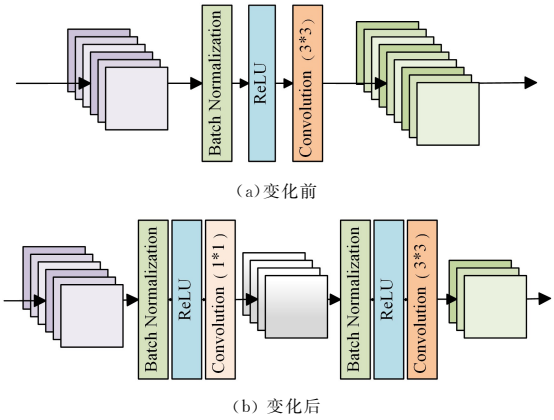


图 2 加入瓶颈层前后局部结构变化图

Fig. 2 Local structural changes chart of before and after adding the bottleneck layer

(2)压缩结构(Compression)

Transition 层除了连接两个相邻的 Denseblock 层，它还可以降低特征图的大小，为了使模型更紧凑，可以减少转换时的特征映射数量层。

假设一个稠密块包含 m 个特征图，穿过过渡层生成 θ_m 个输出特征图，其中 $\theta(0<\theta\leqslant 1)$ 为压缩因子。当 θ 取值为 1 时，穿过过渡层的特征图数量保持不变。将 $\theta<1$ 的 DenseNet 称为 DenseNet-C，在实验中，设置 $\theta=0.5$。

当 Denseblock 中使用瓶颈层，并且过渡层中的 θ 值小于 1 时，将该模型称为 DenseNet-BC。

2.2 注意力机制

注意力机制同神经网络一样都是受到仿生学的启发，它模仿了人类视觉注意力机制。人类视觉通过观察全局图像，选取一些局部重点关注区域，然后对这些区域投入更多注意力来获取更多的细节信息，抑制其他无用信息。

在图像分类任务中，我们多使用柔性注意力机制。本文采用柔性注意力机制中的通道域注意力机制。基于通道域注意力机制的 SENet[18] 网络通过得到每个通道与重要信息之间的关联度，产生基于通道域的注意力机制。在通道域注意力机制中，Squeeze 和 Excitation 是两个非常关键的操作。Squeeze 操作是顺着空间维度进行特征压缩，将每个二维的特征通道变成一个实数。Excitation 操作通过参数 w 来为每个特征通道生成权重，参数 w 被学习用来表示通道间的相关性。

如图 3 所示，我们使用 Global Average Pooling 作为 Squeeze 操作，利用两个 Fully Connected 层组成一个 bottleneck 结构来建模通道间的相关性，并输出与输入特征同样数目的权重。首先将特征维度降低到输入的 1/4($ratio=4$)，然后经过 ReLU 激活后再通过一个 Fully Connected 层使特征维度回升到原来的维度。之后通过一个 Sigmoid 函数获得 0～1 之间归一化的权重。最后通过 Scale 操作将归一化后的权重加权到每个通道的特征上。文献[18]中的压缩比为1/4，即取 $ratio=4$。

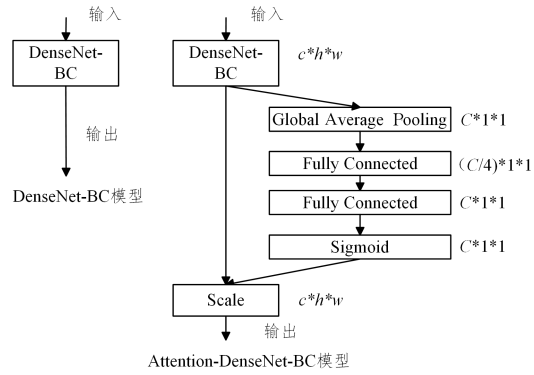


图 3 结合注意力机制的模型结构图

Fig. 3 Model structure diagram with attention mechanism

2.3 基于 Attention 机制的 DenseNet-BC 网络的恶意软件检测模型

在恶意软件分类中，我们选择了适合深度挖掘特征的 DenseNet-BC 网络，它具有特征重用、参数量小的特点，可以有效解决恶意软件特征多样、特征提取复杂的问题，减缓网络梯度消失。

(1)特征重用：在 DenseNet-BC 网络的 Denseblock 中，每一层所学习的特征图会作为输入被直接传递给其后面的所有层。它建立了不同层之间的连接关系，充分利用了恶意软件的特征。

(2)参数量小：以 VGG16 网络举例说明，虽然 DenseNet-BC 的网络层数导致其参数量较大，但其 bottleneck layers 和 Compression 结构可以使网络变窄，大大减少网络的参数量。然而，在 VGG16 网络中，随着网络深度的加深，卷积层的空间复杂度快速提升，每层的空间复杂度是上一层的两倍。DenseNet-BC 网络和本文模型相较于其他模型参数量较小。各模型的参数量大小比较如表 1 所列。

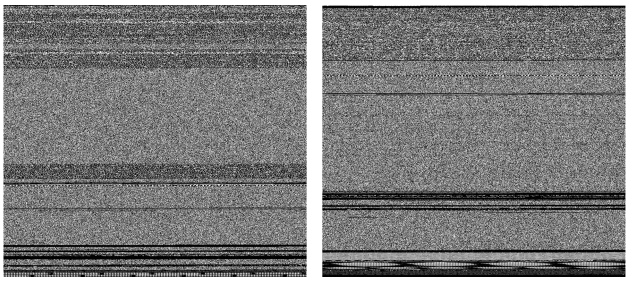
表 1 参数对照表

Table 1 Parameter comparison

Model	Total Params
VGG16	33 698 521
DenseNet	1 878 313
DenseNet-BC	797 863
Attention-DenseNet-BC	1 032 475

为了进一步选择重要特征，提高模型的分类准确率，我们加入了通道域注意力机制，它对各通道赋予不同的权重，选取一部分重点区域，并对这些区域投入更多的注意力来获取更多的细节信息，抑制其他无用信息。在进行恶意软件识别时，部分家族类别恶意软件图像的相似度很高，如图 4 所示。在这种情况下，对于数据量较小的类别，在网络中的学习效果较

差，其测试集的准确率也会较低。使用注意力机制可以更有效地关注纹理特征显著的部分，从而减少对非重要区域的关注。结合 DenseNet-BC 网络和注意力机制，可有效解决恶意软件特征提取复杂以及传统网络中随着网络层数增加引起的梯度消失而造成的准确率下降问题。



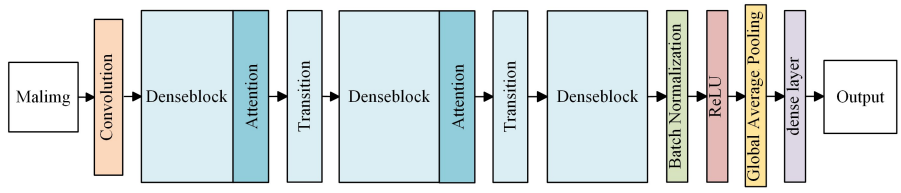
(a)C2LOP. gen!g (b)C2LOP. P

图 4 数据集中的样本图像

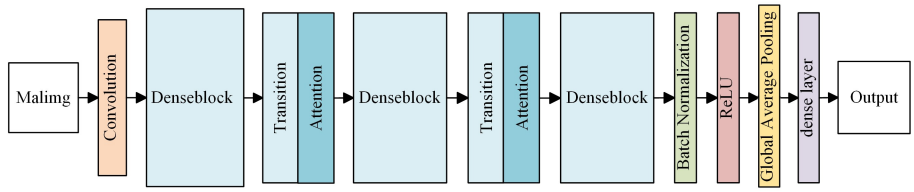
Fig. 4 Sample images in dataset

因为 DenseNet-BC 网络层数较深，且恶意软件数据量很大，只使用一个 Attention 模块难以获得图像中的重要信息，标记分支中出现错误分配权重时难以发现进而修正，因此本文决定使用 2 个 Attention 模块对特征图进行处理。

Attention-DenseNet-BC 模型如图 5 所示，DenseNet-BC 网络为该模型的主干网络，其中有 3 个 Denseblock 模块和 2 个 Transition 层，并且设置了 2 个 Attention 模块。具体实验流程如下：将恶意软件图片作为网络的输入，先经过一层卷积层提取浅层的特征，然后进入 Denseblock，在每个 Denseblock 中有 16 个 bottleneck layers 用作特征提取。之后进入过渡层，并压缩模型参数数量。为了更好地选择重要特征，提高分类准确率，本实验决定将注意力机制在图 5 所示的两个位置分别进行实验，一个是在 Denseblock 模块后加入注意力机制，即图 5 中的位置 1，当特征图通过 Denseblock 层后，为特征图分配权重；另一个是在 Transition 层后加入注意力机制，即图 5 中的位置 2，当特征图通过 Transition 层后，为特征图分配权重。最后一个 Denseblock 后面连接的是全连接层，其赋予每个通道类别标签，然后经过 softmax 完成对恶意软件图片的分类。以上所有模块中的卷积层都有 BN 层和激活函数 ReLU 连接，分类采用全连接。



(a)注意力机制位置 1



(b)注意力机制位置 2

图 5 实验流程图

Fig. 5 Flow chart of experiment

3 实验与结果分析

3.1 标准数据集

本文使用的数据集是一个恶意软件家族公开数据集 Malimg[19]。该恶意软件数据集是最常用于卷积神经网络的恶意软件数据集之一，它包含了来自 25 个不同恶意软件系列的 9339 个恶意软件样本。该数据集可以从 Kaggle 下载。

由于原始数据集图片大小不一，无法放入模型进行训练，因此对原始数据集图片进行处理。对长和宽不等的图片进行填充，使所有图片均变成长宽相等的图片，然后对所有图片进行统一压缩，得到大小均为 64×64 的图片数据。

3.2 评估指标

为了准确、全面地判定本文算法模型泛化性能的优劣，本文采用准确率(Accuracy)、召回率(Recall)、精确率(Precision)和 F1 分数(F1-score)4 种性能度量(performance measure)指标来评价所提模型的恶意软件图像分类性能，评估指标的定义分别如下：

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

$$Recall=\frac{TP}{TP+FN} \tag{5}$$

$$Precision=\frac{TP}{TP+FP} \tag{6}$$

$$F1\text{-}score=2\cdot\frac{Precision\cdot Recall}{Precision+Recall} \tag{7}$$

其中，TP(True Positive)是将正样本分类为正样本的数量值，TN(True Negative)是将负样本分类为负样本的数量值，FP(False Positive)是将负样本分类为正样本的数量值，FN(False Negative)是将正样本分类为负样本的数量值。

3.3 训练结果分析

实验环境为 Centos7 系统，硬件配置如下：CPU 型号为 32 Intel(R) Xeon(R) Silver 4110 CPU @ 2.10GHz，GPU 型号为 NVIDIA Corporation TU104GL [Quadro RTX 5000] (rev a1)，使用 Keras 框架设计实现，其软件配置为 Python3.8，TensorFlow2.3.0 和 Keras2.3.1。实验以恶意软件家族分类任

务为研究背景，对 9339 张实验图片、25 个类别的恶意软件图片进行分类识别。使用基于注意力机制的 DenseNet-BC 模型进行实验，实验过程中设置 200 个迭代周期，模型中的增长率(growth rate)设置为 12，压缩系数(compression)设置为 0.5，初始学习率设置为 0.01，学习率的衰减采用阶梯型衰减，当迭代周期达到 40 时，学习率下降为当前的 1/10，当迭代周期为 80 时，学习率继续下降为当前的 1/10。在反向传递阶段使用的优化方法为随机梯度下降法，其中动量参数(momentum)设置为 0.9，批量大小(batch size)设置为 64。

(1)不同模型的分类效果对比分析

为了验证引入注意力机制的有效性，本文将所提模型 Attention-DenseNet-BC 与 VGGNet[7]，AlexNet[5]，DenseNet，DenseNet-B，DenseNet-C，DenseNet-BC[10]，SE-Inception，SE-Resnet[12]恶意软件分类网络模型进行了对比实验。本次实验中本文模型的注意力机制在 Denseblock 模块后(即图 5 中的位置 1)。为了保证实验的公平性，所有实验都在相同的实验环境中进行，训练集与测试集之比为 6∶4，得到的测试集结果如表 2 所列。

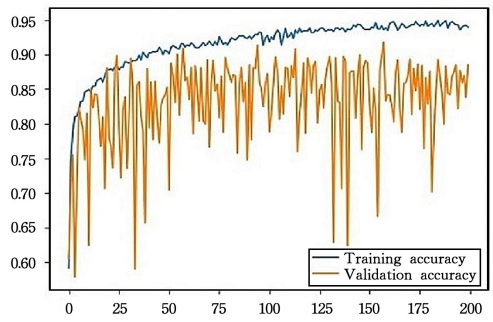
表 2 各模型的分类结果评估

Table 2 Evaluation of classification results of each model

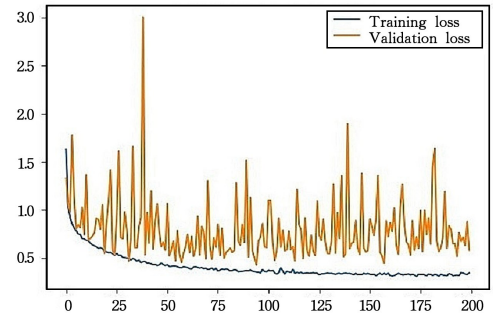
Model	*Accuracy*	*Recall*	*Precision*	*F1-score*
VGG16	0.9098	0.9352	0.9260	0.9305
AlexNet8	0.8672	0.8672	0.8956	0.8811
DenseNet	0.8993	0.8993	0.8947	0.8927
DenseNet-B	0.8990	0.8990	0.8938	0.8969
DenseNet-C	0.8934	0.8934	0.9008	0.8970
DenseNet-BC	0.9036	0.9036	0.9144	0.9089
SE-Inception	0.8851	0.8851	0.8827	0.8839
SE-Resnet	0.9004	0.9004	0.9010	0.9007
our paper	**0.9190**	0.9190	0.9126	0.9157

实验结果表明，在恶意代码检测中，相比其他模型，本文模型达到了较高的准确率，其中准确率、召回率、精确率、F1 分数分别达到 0.9190，0.9190，0.9126，0.9157，其准确率为所有模型中最优。其他模型的各项指标都不及本文模型，相比各项指标都很相近的 VGG16 模型，尽管本文模型仅在准确率上略有提升，但是 VGG16 的参数量要远远大于本文模型的参数量。我们在 2.1.2 节详细说明了 DenseNet-BC 网络的 bottleneck layers 和 Compression 结构可以大大减少该网络的参数量，在 2.3 节说明了 Attention-DenseNet-BC 网络的参数量远小于 VGG16 网络。参数量越高，其模型空间复杂度越高，进而会影响网络性能。虽然 VGG16 网络在 4 项评价指标中的表现并不逊于本模型，但是综合考虑来看，本文模型的空间复杂度低，准确率也表现最优，因此在恶意软件分类应用方面优于 VGG16 模型。

实验中的准确率和损失函数图如图 6 所示。由图可知，注意力机制在位置 1 处的 Attention-DenseNet-BC 模型泛化能力较差，我们通过实验进一步调整参数以改善当前结果。



(a)准确率随 epoch 变化的曲线图



(b)损失函数随 epoch 变化的曲线图

图 6 准确率和损失函数随 epoch 变化的曲线图

Fig. 6 Graph of accuracy and loss function changing with epoch

(2)参数调整对模型分类效果的分析

本实验决定对模型中的部分参数进行调整，以寻找模型的最优参数。

我们对模型中训练集和测试集的不同分割率进行了测试训练，当训练集占整个数据集的 10%时，得到的准确率为 92.61%，优于训练集占比 40%的模型。

在调整训练集占比的基础上，我们对注意力机制在 DenseNet-BC 网络中的位置进行调整，将注意力机制原来所在的位置 1 调整到位置 2(见图 5)，得到的比对结果如表 3 所列。

表 3 注意力机制在不同位置处的测试集性能

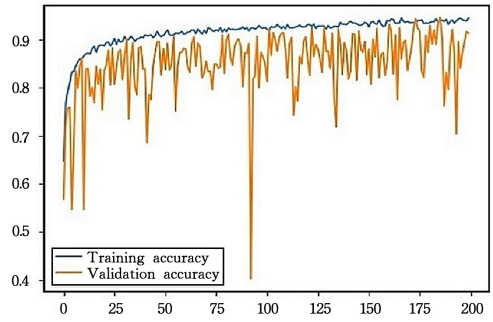
Table 3 Test set performance at different attention mechanism locations

位置	*Accuracy*/%
位置 1	92.61
位置 2	94.64

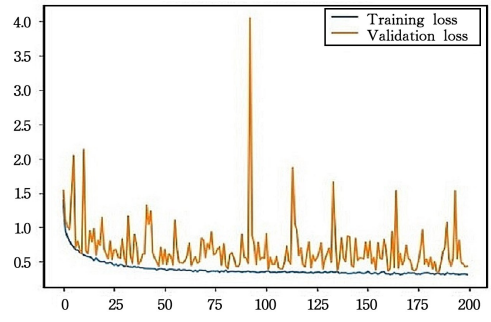
由表 3 可知，在位置 2，即在模型中的两个 Transition 层后面加入注意力机制可以达到更高的准确率。我们认为原因是每张特征图经过 Transition 层中压缩结构的压缩后，特征更为显著，注意力机制为特征显著的部分赋予更高的权重。然而，在只经过 Denseblock 层时，通道数减少，在特征图并未压缩的情况下加入注意力机制，则其选择增强的通道权重差值很小，在经过的网络层数相同的情况下，分类精度必然不如在位置 2 加入注意力机制的效果好。

在调整加入注意力机制的位置的基础上，我们又对优化器进行了重新选择，也调整了注意力机制中的 *ratio* 参数，最后得到准确率最高的模型仍然是上述模型。该模型的优化器

为随机梯度下降，*ratio*＝4，加入注意力机制的位置为 2 个 Transition 层后，其准确率可达到 94.64%。注意力机制在位置 2 处的 Attention-DenseNet-BC 模型的准确率和损失函数图如图 7 所示。该模型相较于注意力机制在位置 1 处的 Attention-DenseNet-BC 模型（见图 6），其泛化能力有所提升，在加深网络的同时，有效解决了卷积神经网络由于深度增加产生梯度消失而造成的准确率下降的问题。



（a）准确率随 epoch 变化的曲线图



（b）损失函数随 epoch 变化的曲线图

图 7　调参后准确率和损失函数随 epoch 变化的曲线图

Fig. 7　Graph of accuracy and loss function changing with epoch after adjusting parameters

结束语　恶意软件检测对网络安全具有非常重要的实际意义。本文使用了结合注意力机制的 DenseNet-BC 网络对恶意软件进行检测，该模型侧重特征重用且加入了视觉注意力，解决了处理图片时特征挖掘不充分，以及卷积神经网络随着网络层数增加引起梯度消失进而导致网络性能退化、准确率较低的问题，是一种较好的恶意软件检测分类模型。特征图通过稠密的内部网络和注意力机制后，特征显著的部分被赋予更高的权重，使得该模型在恶意软件家族分类实验中取得了较好的结果。同时，考虑到各类别的实验数据不平衡也会影响分类准确率，下一步决定加入对数据的均衡处理。

参 考 文 献

[1] CNNIC. The 46th China Statistical Report on Internet Development [EB/OL]. (2020-09-29). http://www.gov.cn/xinwen/2020-09/29/content_5548175.htm.

[2] CNCERT. Analysis Report of China's Internet Network Security Monitoring Data in the First Half of 2020[EB/OL]. (2020-09-26). http://www.cac.gov.cn/2020-09/26/c_160268285 4845452.htm.

[3] ZHANG C, GUO Y, LI M. A review of development and application of artificial neural network models [J/OL]. Computer Engineering and Applications. https://kns-cnki-net.webvpn.bjmu.edu.cn/kcms/detail/11.2127.TP.20210402.1348.004.html.

[4] WANG G D, LU T L, YIN H R, et al. Malicious Code Family Detection Technology Based on CNN-BILSTM[J]. Computer Engineering and Applications, 2020, 56(24): 72-77.

[5] LONG T Y, WAN L, DING H W. Research on the Application of Autocoding Network in Javascript Malicious Code Detection [J]. Computer Science and Exploration, 2019, 13(12): 2073-2084.

[6] HAO J W, LUO S L, ZHANG H Q, et al. Android malicious APP multi-view family classification method [J/OL]. Journal of Beijing University of Aeronautics and Astronautics. https://doi-org-443.webvpn.bjmu.edu.cn/10.13700/j.bh.1001-5965.2020.0658.

[7] LI Y, LUO S L, HAO J W, et al. Malware family classification method based on abstract assembly instructions[J/OL]. Journal of Beijing University of Aeronautics and Astronautics. https://doi-org-443.webvpn.bjmu.edu.cn/10.13700/j.bh.1001-5965.2020.0568.

[8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// the Proceedings of the 25th International Conference on Neural Information Processing Systems. 2012: 1097-1105.

[9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-Based Learning Applied to Document Recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[10] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014(7): 21-34.

[11] SZEGEDY C, LIU W, JIA Y Q, et al. Going Deeper with Convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.

[12] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the Inception Architecture for Computer Vision[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2818-2826.

[13] HE K, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[14] LONG J, SHELHAMER E, DARRELL T, et al. Fully Convolutional Networks for Semantic Segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[15] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: A Search Space Odyssey[C]// IEEE Transactions on Neural Networks and Learning Systems. 2017: 2222-2232.

[16] CHO K, MERRIENBOER B V, GULCEHRE C, et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation[C]// Proceedings of the 2014 Confe-

rence on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2014: 1724-1734.
[17] HUANG G, LIU Z, WEINBERGER K Q, et al. Densely Connected Convolutional Networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
[18] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[C]// IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019: 2011-2023.
[19] BHODIA N, PRAJAPATI P, TROIA F D, et al. Transfer Learning for Image-Based Malware Classification[C]// International Conference on International Workshop on Formal Methods for Security Engineering. 2019.

LI Yi-meng, born in 1997, postgraduate. Her main research interests include network information defense and so on.

LI Cheng-hai, born in 1966, Ph. D, professor. His main research interests include evidence theory, embedded systems, and network security.