

题目：基于CNN的恶意代码检测

组名：暴杀恶意代码组

组长：韦文皞

组员：陈裕中、陈明灿、刘宣、张梦沛、吕浩程

时间：2021年8月24日星期二

指导教师：王勇

1.项目概述

1.1.目的

通过深度学习中的CNN模型来进行恶意代码的静态检测，能够对恶意代码进行一个多分类，确定其恶意代码的类型。

1.2.项目背景

恶意代码攻击已经成为了当今互联网的重要威胁之一，当今现存的恶意代码数据庞大，而且特征多样。所以为了更好的识别出恶意代码以及检测出恶意代码的类型，就需要对恶意代码进行特征提取以及掌握恶意代码的行为，那么卷积神经网络也就是CNN模型对于特征提取工作具有巨大作用，因此通过CNN的方式来识别以及检测恶意代码，可以有效的减轻人工分析的负担，提高互联网的安全性。

1.3.主要功能

基于CNN模型，能够检测出恶意代码以及恶意代码的类型。

1.4.标准

能够正确识别恶意代码的类型，准确率需要达到75%及以上。

2.项目准备

2.1.编程语言

python3

2.2.模型框架

tensorflow, pytroch

2.3.恶意代码管理框架

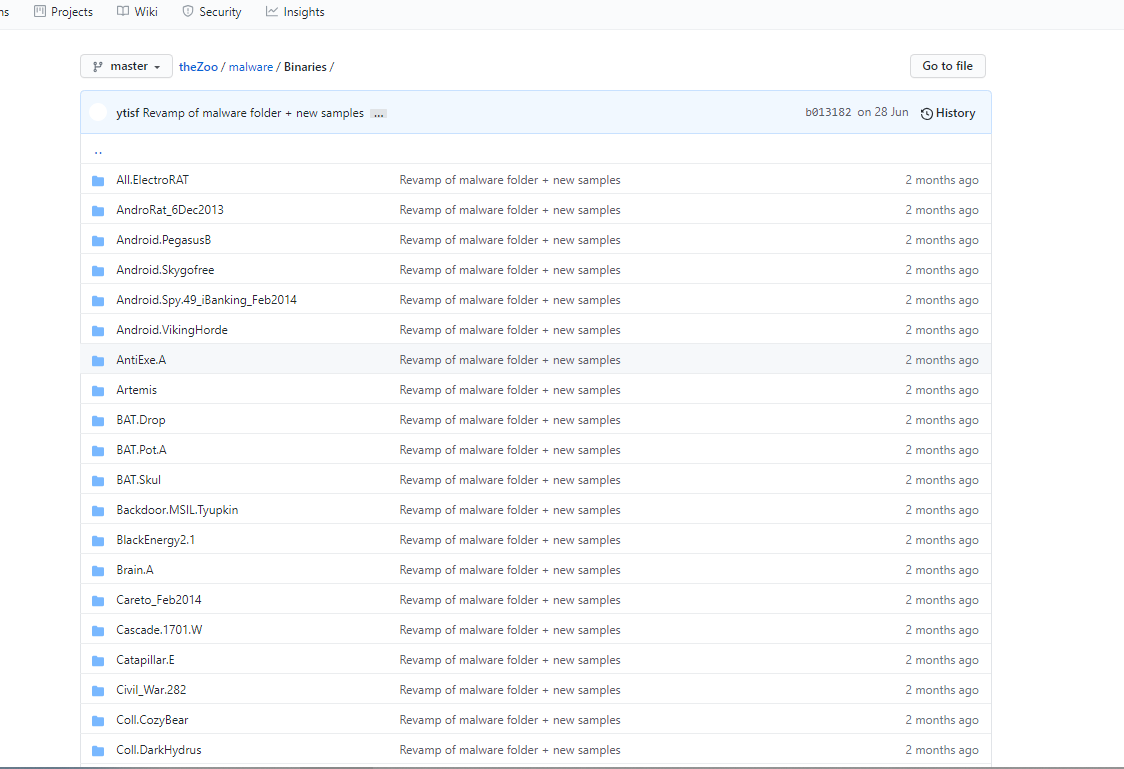
viper

3.项目流程

3.1.数据集

数据集的主要来源

<https://github.com/ytisf/theZoo/tree/master/malware/Binaries>



后续将对恶意代码的存储格式主要使用二进制的存储方式，并将其存储在高斯数据库中，对测试数据集以及训练数据集进行统一管理。(也考虑额外添加别的数据集)

3.2.数据预处理

选取合适的减噪手段以及对数据进行归一化处理，同时根据二进制使用从cv2将二进制数据集转化为一张一张灰度图。

3.3.模型搭建

在进行预处理后恶意代码转化为了灰度图，那么工作也就转换成了基于机器学习的图像识别。了解机器学习中常用的模型，并尝试使用上述框架进行搭建如VGG,ALEX等模型。然后进行训练和测试。

3.4.恶意代码管理框架

了解并使用viper来进行管理，并将已有的模块集成到viper中，后续也可将自己编写的模块也集成到viper中进行统一管理。

3.5.确认访问接口

集成完后确认Restful API可以正常访问，可以直接通过viper来对集成模块来进行使用。