人工智慧導論 HW2

M11107323 羅睿成　　M11107308 游淮君　　M11107322 鍾耀揚　　M11152025 陳彥合

April 15, 2023

1 Introduction

2 Related Work

2.1 Malware analysis

2.1.1 Malware static analysis

2.1.2 Malware dynamic analysis

2.2 Malware visualisation

2.3 Neural network

With the increasing of computational capability of hardware, modern neural network architecture contains huge amount of parameters. Neural network have achieved tremendous performance on regression, classification and generation problem. In this section we introduce two kind of architecture. One isConvolutional Neural Network, and another one is so called Siamese Neural Network.

2.3.1 Convolutional neural network

Convolutional Neural Network (CNN) is a type of deep neural network commonly used for image recognition and processing. CNNs consist of several layers, including convolutional layers, pooling layers, and fully connected layers. In the convolutional layers, the network applies a set of learnable filters to the input image, producing a set of feature maps that capture different patterns and edges in the image. The pooling layers then downsample the feature maps, reducing their size while retaining their important features. Finally, the fully connected layers take the output of the previous layers and use them to classify the image. Figure.1 shows the general architecture of CNN model.

2.3.2 Siamese neural network

A Siamese neural network is a type of neural network architecture that is commonly used for tasks related to similarity or distance-based learning. It consist two or more identical subnetworks, each of them process two or more input data points and output a vector representation for each of them. The vector representations are then compared using a distance metric to determine the similarity between the input data points.

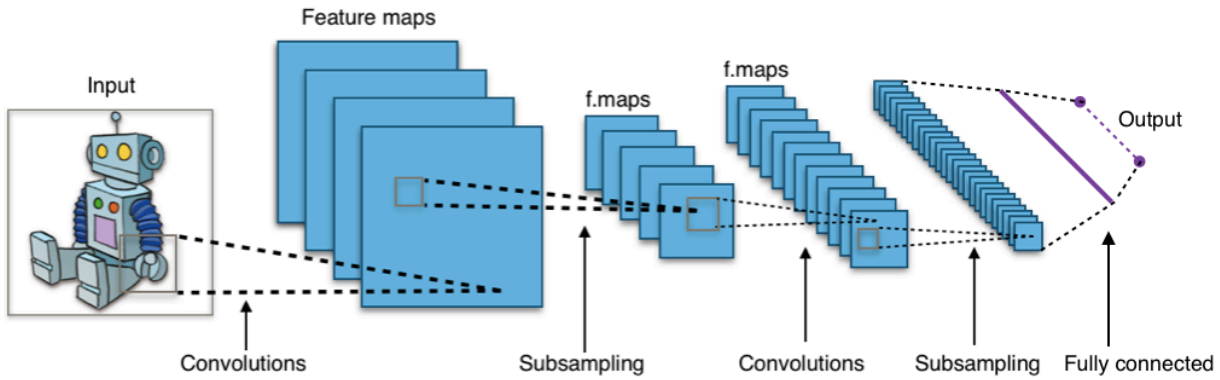

Figure 1: CNN architecture

Siamese networks are widely used in tasks such as image or text matching. They are particularly useful when dealing with small datasets or datasets with high variability, as they can learn to recognize patterns and similarities without requiring a large amount of data. Figure 2. demonstrate the Siamese neural network.

3 Proposed Idea

With the advancement of automated malware generationa and obfuscation, traditional detection to malware are gradually losing their effectiveness or applicability over time. In [**?**] *Mingdong et al.* proposed an idea to extract API calls using dynamic analysis and then map it into feature image based on colour mapping rules. They trained a CNN model to classify 9 class of malware families. and 1000 variants. Figure 3. shows the overview of dynamic API call malware detection.

3.1 Malware API extraction

3.2 Colour mapping rules

3.3 New idea

To collect huge dataset for malware API call is not easy. To reduce the dataset demand for training neural network, we proposed and new idea that is to switch CNN into **Siamese neural network for few-shot training**.

4 Expected Result

Due to the challenges of collecting large and diverse datasets on rapid evolution of Malware, traditional Convolution Neural Network(CNNs) may struggle to fully show their strenths.As mention in our Related Work ,Siamese neural network have been widely using in image or text matching,and are known for their adaptability to small or High variability datasets.Therefore,based on our New idea we'll attempt to utilize Siamese neural networks with few shot learing to achieve a better accuracy then CNNs.

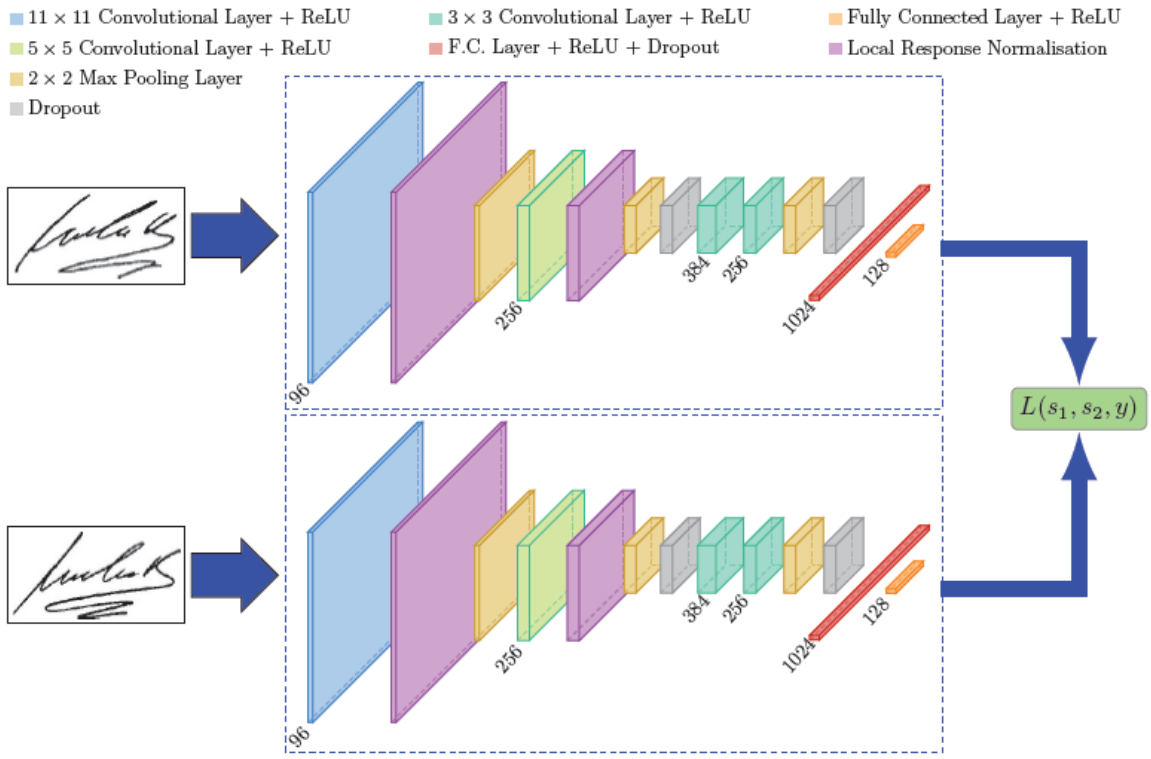

Figure 2: CNN architecture

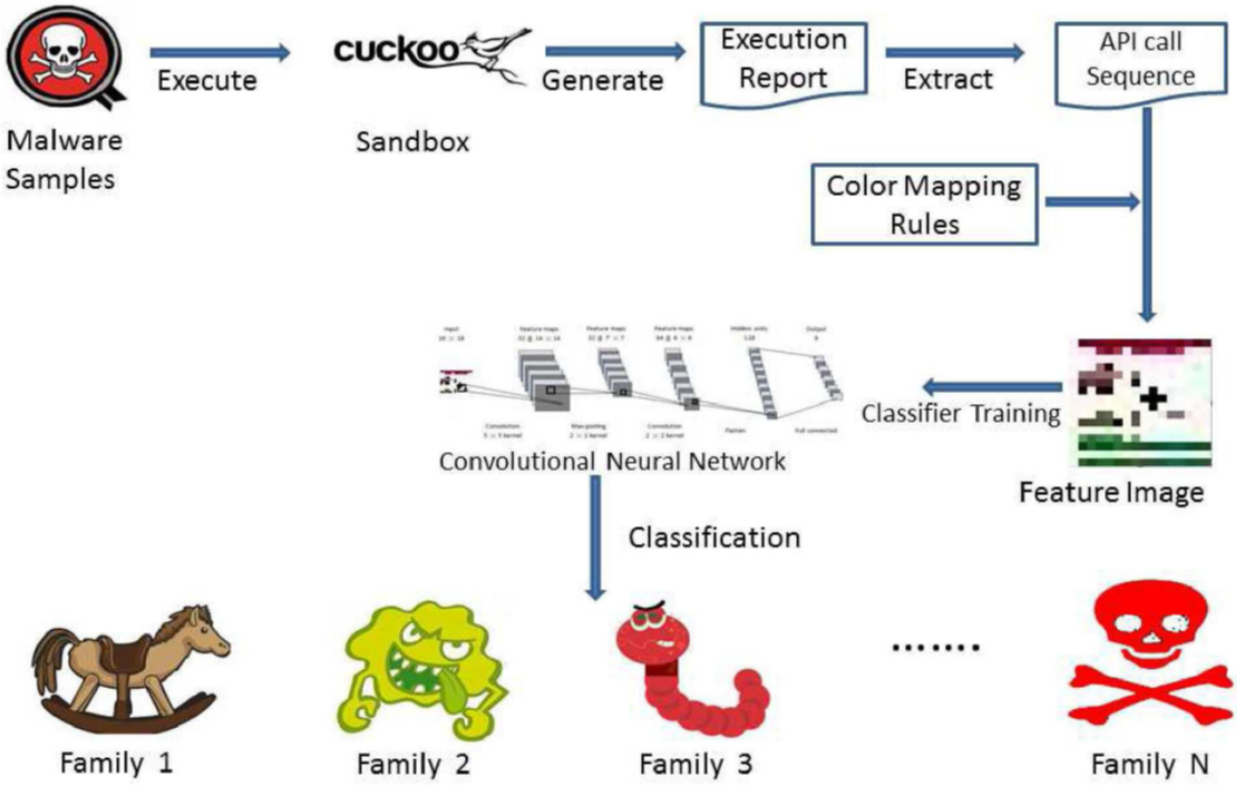

Figure 3: CNN architecture

References

[1] Mingdong Tang, Quan Qain.: 'Dynamic API call sequence visualisation for malware classification'. IET information security, 2019: 1751-8709.