Advances in Social Science, Education and Humanities Research, volume 638
Proceedings of the 2021 International Conference on Public Art and Human Development (ICPAHD 2021)

The Application of Convolutional Neural Network in Malware Images Classification

Shiyu Wang[1a *], Zehao Li[1b], Xiaotian Zhao[1c]

[a] *Dalian University of Science and Technology, LiaoNing, Dalian, CN,116052*
[b] *Rensselaer Polytechnic Institute, Troy, NY, US, 12180*
[c] *University of Minnesota Twin City, Minneapolis, MN, 55455*
**Corresponding author. Email: 18941187296@163.com*

ABSTRACT
Malicious software is a fundamental challenge to information security, which can hijack browsers, force software installation, automatically pop-up ads on web pages, and even support intelligence gathering and destructive cyberattacks. There are always various malicious software and malicious programs on both computers and mobile phones, which have a bad influence on society and people's life. It is important to find a way to recognize them and clean them up. Most new malware is a variant of known malware samples, which can be divided into different types so that each of the same types of malwares has highly similar behaviour characteristics. Therefore, these shared characteristics between malicious samples belonging to the same type can be used to detect and classify unknown programs. Deep learning has achieved good effect in malware classification assignment that converts malware into grayscale images and facilitated the improvement of classification tasks, because models using deep learning convolutional neural network (CNN) can embrace images as input simply. Based on these conditions and combined with the related documents, this paper analyses the nature and mechanism of CNN to classify the current malwares and proposes some possible prospects of it. Finally, it is concluded that compared with ordinary machine learning, the convolutional neural network in malware images classification improves the accuracy of malware classification and reduces the time needed for classification.

Keywords : *Convolutional neural network, Malicious Software classification, Deep learning, Machine learning*

1. INTRODUCTION

The convolutional neural network is a direct end-to-end learning method based on "input-output.", it takes one-time picture pixel facts as input to retain all data of the input photo to the most extent. Feature extraction and high-level abstraction are performed through convolution operation, and mannequin output is the recognition result. Parameters in the model can be trained by back propagation and gradient descent methods. CNN can automatically learn the original data features after training and can extract the learned features and classify them.

The lower sampling and convolution layers of the convolutional neural network are not fully connected, and the mapping between layers is also non-linear. The parameters that need to be trained can be reduced through the three aspects of the local perception field, weight sharing, and pooling. In recent years, CNN have been widely facilitated in image identify, natural language processing, speech recognition and have shown excellent development potential. The visualization of malicious software as grayscale image is an outstanding result of malicious software classification. Nataraj et al. [1] first proposed a method to visualize malware as grayscale images. Nataraj et al. mapped malicious code binaries into images, taking advantage of the Gabor filter's multi-scale and multi-direction characteristics.

This paper studies the classification of malicious images by CNN and expounds and summarizes its essence and methods. With the method of CNN, the malware will be categorized precisely and harmful influences of malware on society and people are expected to be reduced as well.

2. THE DEVELOPMENT AND FEATURE OF CNN

2.1. The development of CNN

2.1.1 Preliminary concept

CNN is a typical deep learning Network architecture inspired by the natural visual cognitive mechanism of biology. In 1959, Hubel & Wiesel [2] discovered that the visual cortex is hierarchical in the information processing of the visual system. For example, when the human eye gazes at a balloon, it starts with the initial signal intake (pixels) the student needs, followed by preliminary processing (the person's various cells draw a rough outline), abstraction (the brain judges the shape of the object in the brain), And deeper abstractions (the brain makes the final decision that the object is a balloon).

2.1.2 The prototype of contemporary convolutional neural network.

In 1998, LeCun et al. [2][3] published the paper lenet-5, after which they designed a multi-layer artificial neural network named lenet-5, which can classify handwritten numbers.

Like other neural networks, LENet-5 can also be trained using the backPropagation [4] algorithm. BP algorithm is applied to the training of this neural network structure, and the modern convolutional neural network prototype is formed.

2.1.3 The success of Alexnet

Since 2006, many techniques have been designed to overcome the issue of coaching deep CNN. CNN can gain the wonderful illustration of the original image, which allows CNN to discover the visible guidelines without delay from the authentic pixels with little pre-processing. However, due to the lack of large-scale statistics and the computer's computational power, lenET-5 used to be no longer perfect for fixing complicated problems. Krizhevsky et al., the most famous, proposed a classical CNN structure and made a breakthrough in image recognition tasks. The overall framework for its approach is called AlexNet [5], similar to Lenet-5. However, it is a little bit more hierarchical. The nonlinear activation function ReLu [6] and Dropout [7] methods are used to achieve excellent results. Until 2012, in the Imagenet Image Recognition contest, Alexnet, stated in The Hinton group's paper "ImageNet Classification with Deep CNN," brought new Deep shape and dropout methods. It raised the error rate from more than 25% to 15% and revolutionized the field of image recognition. After the success of AlexNet, researchers put ahead different enhancement methods, amongst which the most well-known are ZDNet [8], VGGNet [9], GoogleNet [10], and ResNet [11]. From the standpoint of structure, one course of CNN's improvement is to have greater layers. ResNet, the ILSVRC 2015 champion, is more than 20 times that of AlexNet and more than eight times that of VGGNet.

2.2 Convolutional Neural Network processing methodology

Before the Convolutional Neural Network (CNN), it was a great challenge for people to process an image using artificial intelligence. The previous image processing often encountered two problems. The first is that the image usually contains more data to process. For instance, one 1920*1080 resolution picture could have 2,073,600 pixels, and each pixel has R, G, B three parameters. Then processing a single image should cost us to process 6,220,800 parameters. If wanting to process millions of photos or higher resolution photos, such a mountain of data processing is an excessive use of resources. So, the efficiency of processing the image is very low. The second is the inability to retain data information when digitizing images fully. If changing the object's position in the picture, the result would be different using the traditional picture digitization method. This would result in low accuracy of image processing.

With the research on vision principles, people build a multi-layer neural network by imitating the principles of human vision. The lower level recognizes the primary image features, following several lower-level features from a higher level of features, then through the combination of multiple levels, finally make a classification at the top level. A typical CNN consists of three parts: convolutional layer, pooling layer, and fully connected layer. The convolutional layer is responsible for extracting local features in the image.

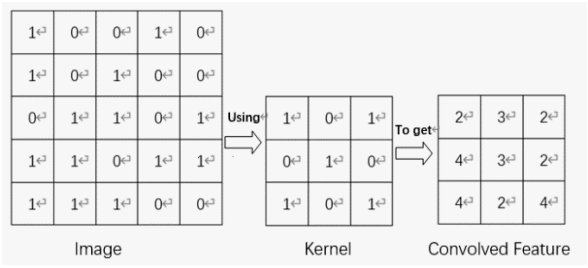

Figure 1: An illustration of the convolutional layer calculation process using a convolution kernel to scan the complete picture.

The pooling layer is simply down-sampling, which can significantly reduce the dimensionality of the data. This is because the image is still huge even after the convolution process is done (because the convolution kernel is relatively small), so down sampling is performed to reduce the data dimension. Generally, there are two sorts of pooling: max pooling and common pooling. Max pooling returns the maximum value of the

image portion, and common pooling returns the common value of the photo portion.

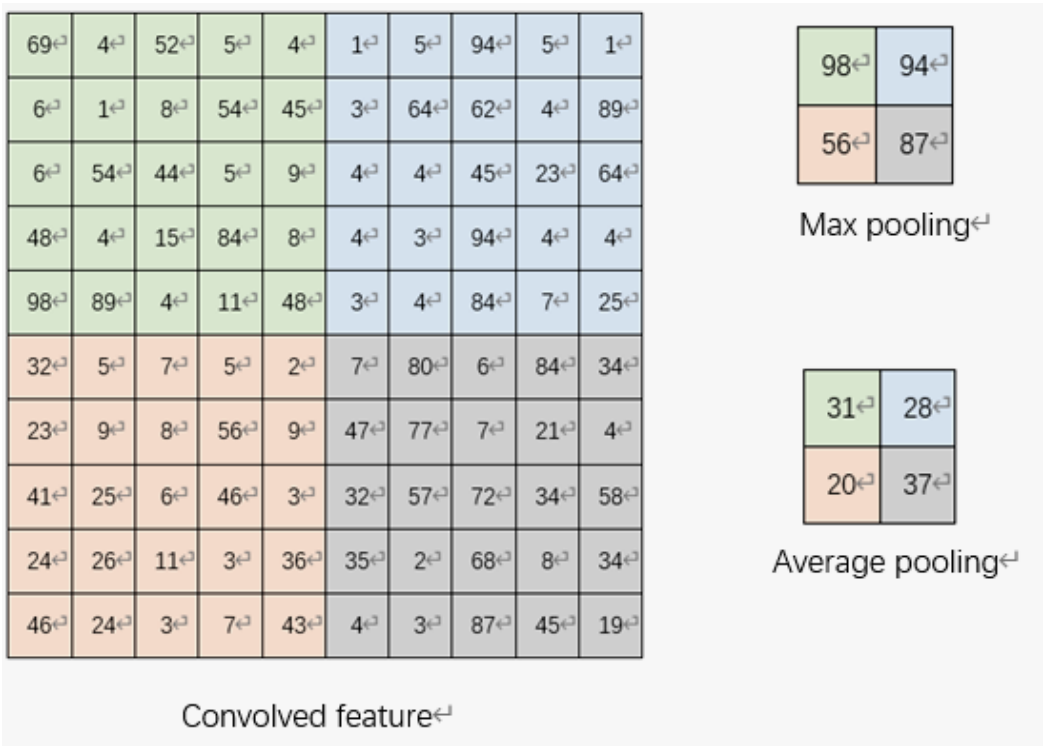

Figure 2: An illustration of the pooling layer calculation process to down sampling the convolved feature.

The final step of CNN is through the fully connected layer. It is responsible for converting the two-dimensional feature map output by the convolution into a one-dimensional vector, thereby achieving an end-to-end learning process. It integrates the feature representation into one value, and its advantage lies in lowing the have an effect of characteristic positions on the classification outcomes and enhancing the robustness of the entire network.

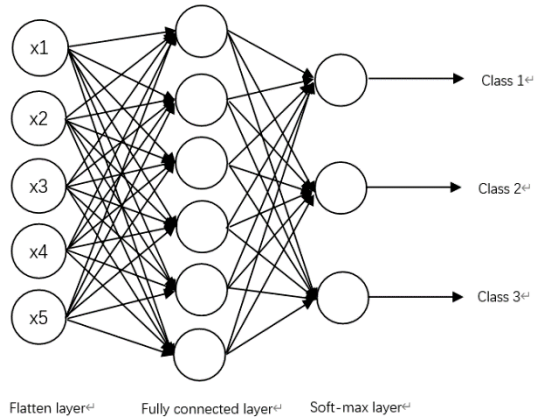

Figure 3: An illustration of the fully connected layer classification process

CNN is very good at processing images and currently has many mature applications. For example, image classification, retrieval, target location detection, target segmentation, face recognition, etc.

2.3 The characteristics of Convolutional Neural Network

CNN have three significant characteristics: local connectivity, parameter sharing, and down sampling. These characteristics make CNN a suitable method for image processing.

In a regular multilayer perceptron, hidden layer nodes are absolutely related to every pixel of an image. In the CNN, every hidden layer node is solely related to a ample enough local pixel of the image, drastically decreasing the weight parameters that want to be trained. In the full connected layer classification process, CNN turns the full connection into a local connection through convolution operation, which shows the local connectivity feature. Next important characteristic is parameter sharing. When processing convolutional layer calculation, one of a kind snapshot or the equal photograph share a single convolution kernel. The equal aspects may additionally show up in the equal image, and the shared convolution kernel can further reduce the weight parameters. Moreover, In CNN, each time the original image is convolved, a down-sampling process is used to reduce the size of the image. This pooling process can make these statistical features have a lower dimension and decrease the amount of calculation. Moreover, it can also reduce the scale of the image and improve the calculation speed. With these

characteristics of CNN, needing to go further discuss the methods used for malware images classification.

3. MALWARE DETECTION METHOD BASED ON CNN

3.1 Malware data set

The malware data set was classified in the Kaggle Machine Learning Challenge, a data analysis competition based on machine learning hosted by Microsoft in 2015. Table 1 shows a total of 10,868 pieces of malware, including Nine different types, about 200GB. This shows that the types of malwares are complex, and therefore, the classification method needs to be put forward.

Table 1. Malware data set

Number	Malware Description	Description
1	Lollipop	Adware
2	VUNDO	Multi-component malware family
3	TRACUR	Trojan
4	OBFUSCATORACY	Method-combination: encryption compression anti-debugging anti- simulation technology
5	KELIHOS V.1	Botnet
6	GATAK	Trojan
7	SIMDA	The most sophisticated malware
8	RAMIT	Powerful botnet functionality
9	KELIHOS V.3	Botnets use polymorphic encryption

3.2 Malware detection method

The CNN is a feed-forward neural network in which the connection pattern between neurons is inspired by the structure of animal visual cortex. This has proved valuable in the analysis of visual images. In order to enhance the accuracy and effectivity of classification of malware detection, put forward a classification of malware detection model is put forward based on CNN, In this model, the first step is to perform the data pre-processing, data normalization processing and to construct an appropriate two-dimensional matrix, then the data will be mapped to the grey image, and then the grey image will be taken as input CNN characteristics, at last the characteristics and the automatic classification will be studied.

4. THE EFFECT OF CNN

In the previous content, we learned the process of accurate image recognition by CNN. Generally, it is the process from the whole to the part and then to the whole. Take an image example. For example, here is a picture of a cat, which we need to identify with CNN. Let us start by asking ourselves why we are so good at judging pictures of cats. Because we receive the comprehensive information of the picture through our eyes, we disassemble it and judge it step by step. CNN works the same way. The cat pictures were disassembled and analysed. Of course, the image case here can be a little more complicated because there are also pixel and gamut issues and the arrangement of the dots. Now we can start with a more straightforward case. For example, we often have pop-ups when we use our computers. The most typical is the advertising of some online games, and this is the popover. In addition to some colourful pictures, there are also some sensitive words. Let us assume that the sensitive word here is' X. ' (Table2) If only the whole image is analysed, judgment accuracy will decrease once the image is displaced. However, if local sampling is carried out, it will be more conducive to data fitting. The blue circle in the figure is used as a sample to make local errors to get the right picture.

Table2: similarity-recognition

1	-1	-1	-1	1
-1	1	-1	1	-1
-1	-1	1	-1	-1
-1	1	-1	1	-1
1	-1	-1	-1	1

Table 3: detail similarity-recognition

-1	-1	-1
1	-1	1
-1	1	-1

Table 4: similarity-recognition calculation

-0.33	1	-0.33
-0.56	-0.78	-0.56
-0.33	1	-0.33

5. DISCUSSION

The growing threat of malware is becoming more and more difficult to ignore. According to Deep Instinct, a cybersecurity company, in 2020, malware accelerated by using 358% overall, and ransomware expanded with the aid of 435% compared in contrast 2019. Moreover, malware threats are on many different platforms, and it is more serious about setting a defence system to protect from malware attacks. In the future, malware image classification using CNN will be deployed into many cybersecurity software to classify different malware. Furthermore, this method has higher scalability to intersect with other methods. We will continue to research to train this method to optimize the output and further verify the generalization ability of our method.

We learned that the process of classification includes the step of input and output. There is undoubtedly a massive improvement from the most primitive simple programming method to training machine self-learning and the accuracy of image judgment.

6. CONCLUSION

In conclusion, this paper demonstrates the development of CNN, introduces the method of malicious software, and investigates the effect of CNN. For future research work, this will be a fruitful area for future work and ensure an exemplary network environment as the Internet will continue to affect our lives. Nevertheless, Recognising the limitations of the current study, an additional uncontrolled factor is that when classifying malicious software as a greyscale image, there may exist some visual-related problems. So, there is still a risk of this technology because minor errors can lead to significant errors. Therefore, building a network security ecology and the recognition of malware is very necessary.

ACKNOWLEDGMENT

When writing this paper, our team has received great support and assistance. Thanks to thank Dr. Vipul Goyal who taught us basic knowledge and had discussion about this topic. And would like to thank our supervisor, for her guidance during writing this paper. Finally, if complete this dissertation that without help from each member of our team is impossible, everyone made significant contribution for completing this dissertation.

REFERENCES

[1] Nataraj L, Karthikeyan Jacob G, Manjunath B(2011). Malware Images: Visualization and Automatic Classification[C]. Proceedings of the 8th International Symposium on Visualization for Cyber Security.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner (1998), "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324.

[3] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hub-bard, and L. D. Jacke (1990) l, "Handwritten digit recognition with a back-propagation network," in Advances in neural information processing systems. Citeseer.

[4] R. Hecht-Nielsen (1989), "Theory of the backpropagation neural network," in International Joint Conference on Neural Networks, pp. 593–605.

[5] A. Krizhevsky(2012), I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105.

[6] V. Nair and G. E. Hinton (2010), "Rectified linear units improve restricted boltzmann machines," in ICML, pp. 807–814.

[7] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov(2012), "Improving neural networks by preventing co-adaptation of feature detectors," arXiv preprint arXiv:1207.0580.

[8] M. D. Zeiler and R. Fergus (2014), "Visualizing and understanding convolutional networks," in ECCV.

[9] K. Simonyan and A. Zisserman (2015), "Very deep convolutional networks for large-scale image recognition," in ICLR.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich (2014), "Going deeper with convolutions," CoRR, vol. abs/1409.4842.

[11] T. Wang, D. Wu, A. Coates, and A. Ng (2012), "End-to-end text recognition with convolutional neural networks," in International Conference on Pattern Recognition (ICPR), pp. 3304–3308.

[12] Simple explanation of convolutional neural network | Deep Learning Tutorial 23 (Tensorflow & Python). https://www.youtube.com/watch?v=zfiSAzpy9NM

[13] M. Kalash, M. Rochan, N. Mohammed, N. D. B. Bruce, Y. Wang, and F. Iqbal, "Malware Classification with Deep Convolutional Neural Networks," (2018 9th) IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp. 1-5, doi: 10.1109/NTMS.2018.8328749.