Dark Web Image Classification Using Quantum Convolutional Neural Network

Ashwini Dalvi
Department of Computer Engineering
Veermata Jijabai Technological Instutute
India
aadalvi_p19@ce.vjti.ac.in

Soham Bhoir
Department of Information Technology
K. J. Somaiya College of Engineering
India
soham.bhoir@somaiya.edu

Irfan Siddavatam
Department of Information Technology
K. J. Somaiya College of Engineering
India
irfansiddavatam@somaiya.edu

S G Bhirud
Department of Computer Engineering
Veermata Jijabai Technological Instutute
India
sgbirud@ce.vjti.ac.in

Abstract – Researchers have investigated the dark web for various purposes and with various approaches. Most of the dark web data investigation focused on analyzing text collected from HTML pages of websites hosted on the dark web. In addition, researchers have documented work on dark web image data analysis for a specific domain, such as identifying and analyzing Child Sexual Abusive Material (CSAM) on the dark web. However, image data from dark web marketplace postings and forums could also be helpful in forensic analysis of the dark web investigation.
The presented work attempts to conduct image classification on classes other than CSAM. Nevertheless, manually scanning thousands of websites from the dark web for visual evidence of criminal activity is time and resource intensive. Therefore, the proposed work presented the use of quantum computing to classify the images using a Quantum Convolutional Neural Network (QCNN). Authors classified dark web images into four categories alcohol, drugs, devices, and cards. The provided dataset used for work discussed in the paper consists of around 1242 images. The image dataset combines an open source dataset and data collected by authors. The paper discussed the implementation of QCNN and offered related performance measures.

Keywords: Dark Web, Marketplace Images, Quantum Circuit, Quantum Computing for Cyber Security, Quantum Convolutional Neural Network.

I. INTRODUCTION

Anonymity networks like Tor enable the hosting of covert internet markets where dark vendors may deal with illegal goods, including narcotics, firearms, and hacking services. Classification and analysis of images collected from the dark web marketplaces will assist in analyzing information shared on these markets. However, existing dark web marketplace investigation mechanisms majorly consider the text, ignoring non-textual objects like images. There are two main obstacles to overcome when analyzing dark web images: (a) moral and legal concerns about the existence of child exploitation imagery in illicit markets and (b) the computational burden of downloading, processing, analyzing, and storing image content.

Also, the open source dark web data is sparse, making it difficult for researchers to train the machine learning model to classify images from different categories.

The typical approach to researching the dark web is to design custom crawlers to collect data from the dark web. However, the volume of crawled data is high; thus, performing image classification with classical machine learning or deep learning with traditional computing facilities will be challenging.

There has been a significant interest in quantum machine learning (QML) by researchers. This is because QML offers the ability to more effectively address the issues associated with massive data and the slow training process in existing conventional machine learning. [8, 10]

Quantum Convolution Neural Networks (QCNNs) handle quantum data to recognize stages of quantum states and create a quantum error-correcting system. Quantum gates with controllable parameters approximate the pooling and convolutional layers. [6, 16]

The authors present a quantum convolutional neural network based on parameterized quantum circuits in the proposed work. To be processed by quantum technology, image data must be encoded into quantum states [8]. The authors use the QCNN model to handle grid-type input, such as photographs.

A parametric rule is used in the proposed model to calculate the analytical gradients of loss functions on quantum circuits and to achieve faster yet more stable convergence [1, 7, 16].

The paper is organized into sections: section II covers related work. Section III discussed the proposed method, followed by section IV on the result and discussion.

II. RELATED WORK

Researchers attempted image analysis with Compass Radius Estimation for Image Classification (CREIC) on the dark web data [3]. The work is one of the few attempts to categorize dark web images into five categories. In other work approaches of perceptual hashing are discussed for dark web image classification. However, the proposed work aims to employ the advantage of quantum computing to classify dark web image data into four categories [3].

Numerous computer science research has been affected by the idea of quantum computation, particularly those in the fields of computational modeling, cryptography theory, and information theory. Information security may benefit from quantum computers, or it may suffer a detrimental effect. Many researchers have thoroughly examined the advantages of quantum computing in cybersecurity. The use of a quantum computer can potentially be advantageous in several domains [11].

Research on the application of many quantum concepts was given.[15], such as Intruder detection systems for healthcare systems may be trained using mechanics and neural networks. The suggested method is tested on the KDD99 dataset.

Researchers investigated how quantum computing may reduce the need for domain-specific security. In [16] it provided that quantum computing-based representations of standard AES and modified AES algorithms to underline that quantum computing will be a viable solution to improve cybersecurity. [11] Quantum cryptography to encrypt communication between sensors and computers to protect cyber-physical systems.[12] A look at the viability of fusing quantum and conventional computers.[13]
A unique hybrid quantum-classical deep learning model for botnet detection using domain generation algorithms (DGA)[14].

Researchers are also examining if applying quantum mechanical concepts to machine learning issues might enhance the outcome. The most recent research results in quantum machine learning were compiled under [12, 14]. A classification with Quantum Convolutional Neural Network strategy was suggested by the authors.

About Quantum Convolutional Neural Network

1. <u>Neurons and Weights</u>

A neural network is a complex function constructed from smaller building components known as neurons. A neuron is often a nonlinear function that translates one or more inputs to a single real number. It is also typically simple, straightforward to compute, and nonlinear. Usually, neurons copy their single output and provide it to other neurons as input. In order to visually depict how the output of one neuron will be utilized as the input to other neurons, we represent neurons as nodes in a graph and draw directed edges between nodes. Also noteworthy is that each edge in our graph frequently has a scalar number called a weight attached to it. According to this theory, each input to a neuron will be multiplied by a separate scalar before being gathered and processed into a single result. In order to train a neural network, the primary goal of the proposed work is to select weights that will cause the network to act in a specific manner.

2. <u>Input Output Structure of neural network</u>

A traditional (real-valued) vector serves as the input to a neural network. According to the network's graph topology, a layer of neurons receives each input vector component multiplied by a distinct weight. Then, the findings are compiled into a new vector, where the i'th component stores the output of the i'th neuron after each neuron in the layer has been assessed. After that, a new layer can use this new vector as an input, and so on. Except for the proposed network's initial and last levels, all other layers are hidden.

3. <u>Feed Forward Neural Network</u>

A feed-forward neural network is a name given to the type of neural network we will be working with (FFNN). This means that information will never hit a cell again as it passes through our brain network [18]. We may call the graph representing our neural network a directed acyclic graph (DAG). Furthermore, no edges will be allowed between neurons in the same neural network layer.

4. <u>Backend</u>

The backend acts as either a simulator or an actual quantum computer, operating quantum circuits and/or pulse schedules and providing results.

5. <u>Shots</u>

1. A single trip through each step of an entire quantum circuit on an IonQ(Trapped ion Quantum Computing), Rigetti, or OQC (Outgoing Quality Control) gate-based QPU(Quantum Processor) is called a "shot."

A Mathematical Approach to Quantum Convolutional Layer

Let X^l be the input and K^l be the Kernel for the layer l of a convolutional neural network,
And f: R $\rightarrow$[0, C] with C > 0 be a non-linear function so that $f(X^{l+1}) := f(X^l * K^l)$ is the output for layer l. The given X^l and K^l are stored in Quantum Random Access Memory (QRAM) [4]; there is a quantum algorithm that, for precision parameters
$\varepsilon > 0$ and $\eta > 0$, creates a quantum state $|f(\overline{X}^{l+1})\rangle$ such that $\| f(\overline{X}^{l+1}) - f(X^{l+1}) \|_\infty \leq 2\varepsilon$ and retrieves classical tensor $\mathcal{X}^{l+1}$ such that for each pixel j. [11, 17]

$$\begin{cases} \left|\chi_j^{l+1} - f\left(\chi_j^{l+1}\right)\right| \leq 2\varepsilon & if\ f\left(\chi_j^{l+1}\right) \geq \eta \\ \chi_j^{l+1} = 0 & if\ f\left(\chi_j^{l+1}\right) < \eta \end{cases}$$

The algorithm has time complexity as

$$O\left(\frac{1}{\varepsilon\eta^2} \cdot \frac{M\sqrt{C}}{\sqrt{E\left(f\left(\overline{X}^{\,l+1}\right)\right)}}\right)$$

O hides the poly-logarithmic in the size of X^l and K^l.

Algorithms Used

1. Forward Pass for QCNN

The quantum analog of a single quantum convolutional layer is implemented in the QCNN forward pass method. To prepare the input for the following layer, it first applies a convolutional function to an input and a kernel, then applies a nonlinear function and performs pooling operations. [11]

2. Quantum Backpropagation Algorithm (Backward Pass)

A widely used algorithm to train feed-forward neural networks is backpropagation [11]. The algorithm required for a quantum convolutional neural network is a quantum backpropagation algorithm. In classical feed-forward neural networks, the classical backpropagation algorithm updates all kernel weights according to the derivative of a given loss function *L*.

The algorithm calculates each element of the gradient tensor $\frac{\partial L}{\partial F^l}$ within additive error $\delta \left\| \frac{\partial L}{\partial F^l} \right\|$, which updates F^l as per the gradient descent update rule.

The time complexity of a single layer *l* for quantum backpropagation is [11]:

$$O\left(\left(\left(\mu(A^l) + \mu\left(\frac{\partial L}{\partial Y^{l+1}}\right)\right)\kappa\left(\frac{\partial L}{\partial F^l}\right) + \left(\mu\left(\frac{\partial L}{\partial Y^{l+1}}\right) + \mu(F^l)\right)\kappa\left(\frac{\partial L}{\partial Y^l}\right)\right)\frac{\log {}^{1}/_{\delta}}{\delta^2}\right)$$

The gradient is calculated as follows:

$$\Delta_{x_0}\ QuantumCircuit(x_0) = QuantumCircuit(\alpha + \beta) - QuantumCircuit(\alpha - \beta)$$

Here x_0 and α represented as a parameter for Quantum Circuit and β a macroscopic shift. Thus the gradient is simply the difference the Quantum Circuit evaluated at $\alpha + \beta$ And $\alpha - \beta$.

III. Proposed Method

The proposed work collected image data with a customized dark web crawler. The crawler was run into the instance to collect the data. The collected images are trained into three categories alcohol, devices, and cards. Figure 1 shows sample images from the crawled data.

Fig 1. Sample images from crawled data

Because of the limitation of usable images from crawled data, the authors used images of drugs from darknet market archives [4]. Figure 2 depicts the dataset creation.

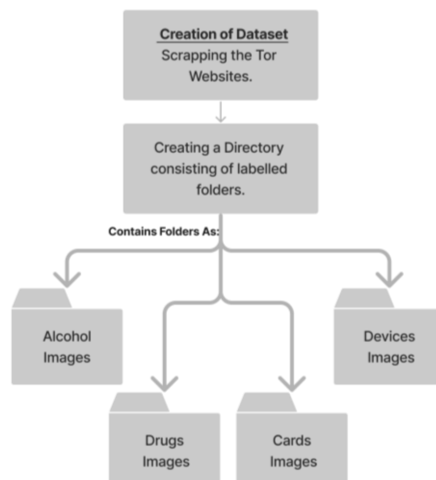

Fig 2 Dataset creation

Figure 3 depicts the proposed methodology for classifying onion services market images with the quantum convolutional neural network.

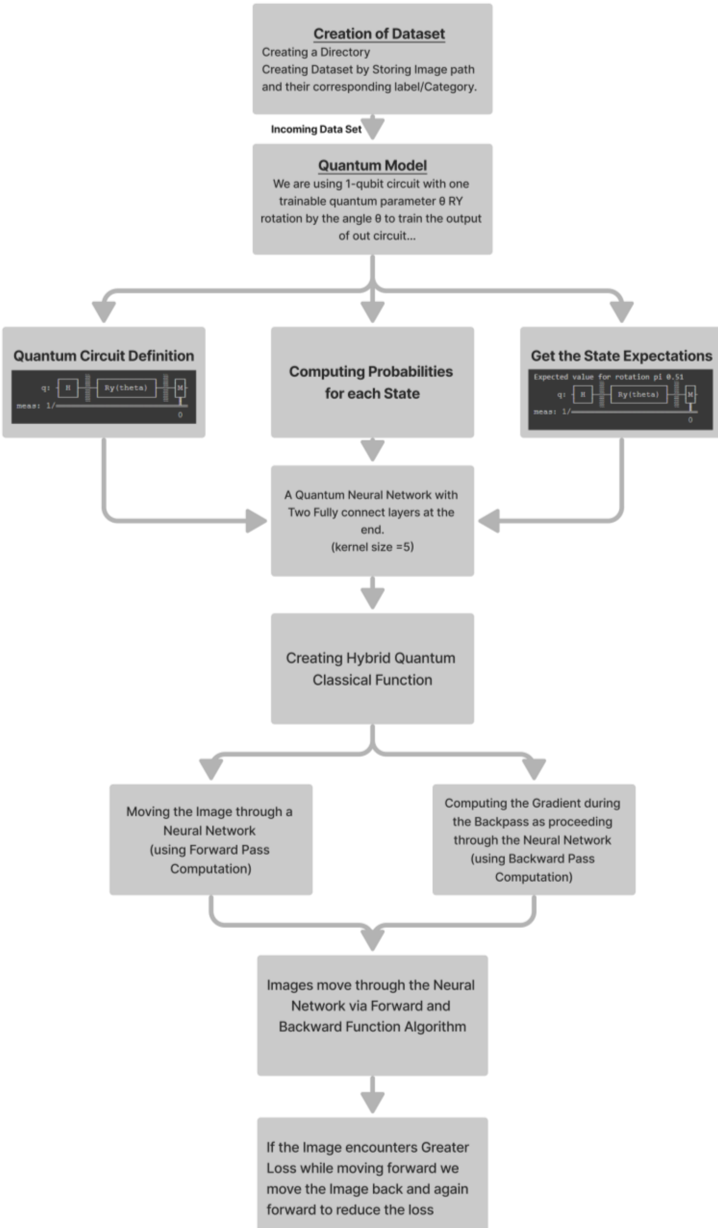

Fig 3. Proposed Method

Quantum Circuit Used

The quantum functions have been organised into a class by the authors. The indications shots is propagated and trainable quantum parameters to employ in our quantum circuit. To keep things straightforward by using a 1-qubit circuit with a single trainable quantum parameter x_0. And use a Ry rotation by the angle x_0 so as to train the output of the parameterized circuit

To measure the output on the basis of Z, authors have calculated the expectation as $\emptyset_z$

$$\emptyset_z = \sum_i^n z_i \, f(z_i)$$

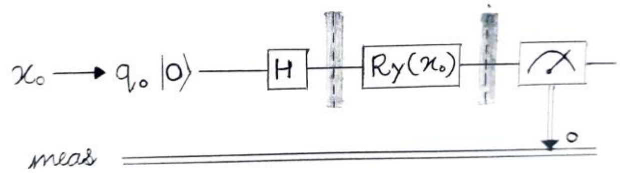

Fig 4 1-qubit quantum circuit

The Neural Network

Authors have use a parameterized quantum circuit to construct a hidden layer for the neural network to produce a quantum-classical neural network. The term "parameterized quantum circuit" refers to a quantum circuit in which each gate's rotation angle is determined by the elements of a classical input vector. The parameterized circuit will be fed with the outputs from the previous layer of the neural network. The quantum circuit's measurement data may then be gathered and utilized [9] as inputs for the layers below.

Pseudo Code for QCNN

The following pseudo-code explains the working of the proposed model, followed by the result and discussion.

Function to create a dataset by storing image path and corresponding label:
Return dictionary with key as image path and value as label index

Class for QuantumCircuit:
constructor taking input as n_qubits, shots, backend:
defining circuit parameters

Function to run the circuit taking parameter as a rotating angle:
counting the result of each Iteration through the backend,
getting the states of each count
computing the probabilities of each state
getting the state expectation

return an array of state expectation

Class for HybridFunction
Static Function for forward pass computation taking context, input, quantum_circuit, and shift as parameters:
getting the shifts from the context
getting the shifts from quantum_circuit

getting the expectations along the Z-axis of rotation

calculating the result as tensors of expectations

storing the input and result for further backward pass computation

Static Function for backward pass computation taking context and gradient_output as parameters:

destructuring input and expectations along Z-axis to save as tensor pair

taking a list of input from the previous gradient

calculating the amount of shift shape

adding to the calculated shift to input_list to shift right

subtracting the calculated shift from input_list to shift left

a loop to append values of gradient after subtracting left and right side expectations

IV. RESULT AND DISCUSSION

Figure 5 shows the performance graph of the proposed model. The data size of 1242 images, and the model performance time is in a few milliseconds. The result shows that quantum-based models significantly reduce training time even with large data sizes.

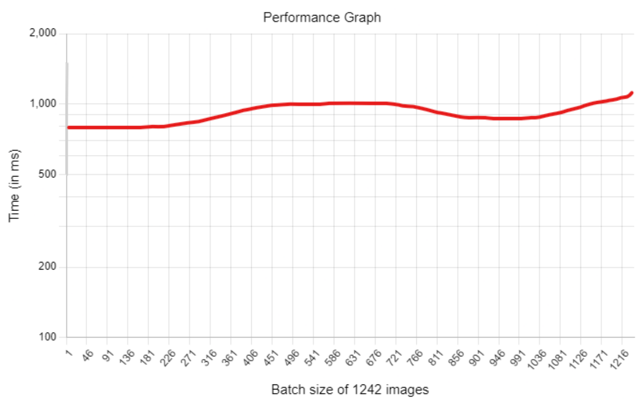

Fig. 5 Performance graph with the batch size of 1242

Machine learning models use epochs to determine which model represents the sample with the lowest amount of error. Before training the neural network, the epoch and batch size must be specified. Figure 6 shows how the negative log-likelihood loss varies over training iterations.

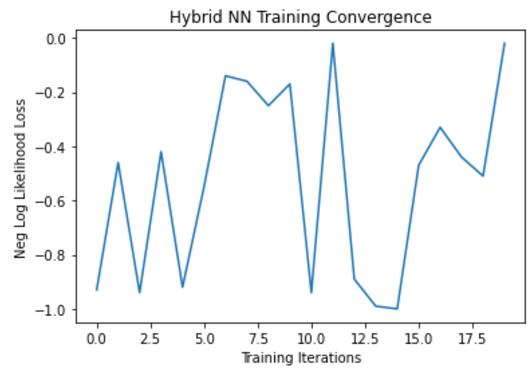

Fig 6 Negative log-likelihood loss v/s training iterations

CONCLUSION

The dark web data interests researchers for a variety of reasons. Typically, the investigation of abusive materials from the dark web is explored by researchers. The proposed work extends the scope of dark web image investigation in two ways: first, collecting a sufficient number of image data sets for work and demonstrating QCNN's effectiveness for dark web image classification. The performance of the proposed model is in milliseconds to classify image data set of 1242 images on the Google colab environment.

Also, the proposed work's novelty is that the image classification is conducted on dark web marketplace data. The combined custom and open data of different classes are used in the presented work.

REFERENCES

[1] Lü, Y., Gao, Q., Lü, J., Ogorzałek, M., & Zheng, J. (2021, July). A Quantum Convolutional Neural Network for Image Classification. In 2021 40th Chinese Control Conference (CCC) (pp. 6329-6334). IEEE.H. Simpson, *Dumb Robots*, 3rd ed., Springfield: UOS Press, 2004, pp.6-9.

[2] Fidalgo, E., Alegre, E., González-Castro, V., & Fernández-Robles, L. (2017, September). Illegal activity categorisation in DarkNet based on image classification using CREIC method. In International Joint Conference SOCO'17-CISIS'17-ICEUTE'17 León, Spain, September 6–8, 2017, Proceeding (pp. 600-609). Springer, Cham.

[3] Biswas, R., Fidalgo, E., & Alegre, E. (2017, December). Recognition of service domains on TOR dark net using perceptual hashing and image classification techniques. In 8th International Conference on Imaging for Crime Detection and Prevention (ICDP 2017) (pp. 7-12). IET.

[4] Oh, S., Choi, J., Kim, J. K., & Kim, J. (2021, January). Quantum convolutional neural network for resource-efficient image classification: A quantum random access memory (QRAM) approach. In 2021 International Conference on Information Networking (ICOIN) (pp. 50-52). IEEE.

[5] Kerenidis, I., Landman, J., & Prakash, A. (2019). Quantum algorithms for Trochun, Y., Stirenko, S., Rokovyi, O., Alienin, O., Pavlov, E., & Gordienko, Y. (2021, September). Hybrid Classic-Quantum Neural Networks for Image Classification. In *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (Vol. 2, pp. 968-972). IEEE.

[6] Wei, S., Chen, Y., Zhou, Z., & Long, G. (2022). A quantum convolutional neural network on NISQ devices. *AAPPS Bulletin*, *32*(1), 1-11.

[7] Zhao, C., & Gao, X. S. (2021). Qdnn: deep neural networks with quantum layers. *Quantum Machine Intelligence*, *3*(1), 1-9.

[8] Hur, T., Kim, L., & Park, D. K. (2022). Quantum convolutional neural network for classical data classification. *Quantum Machine Intelligence*, *4*(1), 1-18.

[9] Li, W., Chu, P. C., Liu, G. Z., Tian, Y. B., Qiu, T. H., & Wang, S. M. (2022). An Image Classification Algorithm Based on Hybrid Quantum Classical Convolutional Neural Network. *Quantum Engineering*, *2022*.

[10] Abohashima, Z., Elhosen, M., Houssein, E.H. and Mohamed, W.M., 2020. Classification with quantum machine learning: A survey. *arXiv preprint arXiv:2006.12270*.

[11] Kerenidis, I., Landman, J., & Prakash, A. (2019). Quantum algorithms for deep convolutional neural networks. *arXiv preprint arXiv:1911.01117*.

[12] Tosh, D., Galindo, O., Kreinovich, V., & Kosheleva, O. (2020, June). Towards security of cyber-physical systems using quantum computing algorithms. In *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)* (pp. 313-320). IEEE.

[13] Ali, A. (2021, January). A Pragmatic Analysis of Pre-and Post-Quantum Cyber Security Scenarios. In *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)* (pp. 686-692). IEEE.

[14] Suryotrisongko, H., & Musashi, Y. (2022). Evaluating hybrid quantum-classical deep learning for cybersecurity botnet DGA detection. *Procedia Computer Science*, *197*, 223-229.

[15] Laxminarayana, N., Mishra, N., Tiwari, P., Garg, S., Behera, B. K., & Farouk, A. (2022). Quantum-Assisted Activation for Supervised Learning in Healthcare-based Intrusion Detection Systems. *IEEE Transactions on Artificial Intelligence*.

[16] Ko, K. K., & Jung, E. S. (2021). Development of cybersecurity technology and algorithm based on quantum computing. *Applied Sciences*, *11*(19), 9085.

[17] Farhi, E., & Neven, H. (2018). Classification with quantum neural networks on near term processors. *arXiv preprint arXiv:1802.06002*.