Quantum Enhanced Filter: QFilter

Parfait Atchade-Adelomou\textsuperscript{1, 2} and Guillermo Alonso-Linaje\textsuperscript{3}

\textsuperscript{1}Engineering Department Research Group on Data Science for the Digital Society La Salle - Universitat Ramon Llull Carre\`{e} de Sant Joan de La Salle, 42 08022 Barcelona (Spain)
\textsuperscript{2}Lighthouse Disruptive Innovation Group, LLC 7 Broadway Terrace, Apt 1 Cambridge MA 02139 Middlesex County, Massachusetts (USA)
\textsuperscript{3}Universidad de Valladolid C/Plaza de Santa Cruz, 8, 47002 Valladolid (Spain)

(Dated: March 2021)

Convolutional Neural Networks (CNN) are used mainly to treat problems with many images characteristic of Deep Learning. In this work, we propose a hybrid image classification model to take advantage of quantum and classical computing. The method will use the potential that convolutional networks have shown in artificial intelligence by replacing classical filters with variational quantum filters. Similarly, this work will compare with other classification methods and system's execution on different servers. The algorithm's quantum feasibility is modelled and tested on Amazon Braket Notebook instances and experimented on the PennyLane's philosophy and framework.

\textbf{KeyWords:} Quantum Computing, Machine Learning, Convolutional Neural network, Quantum Quantum Filter, Artificial Intelligent, Quantum Gradient, QHack2021

I. INTRODUCTION

From a classical point of view, artificial intelligence has appeared as a solution to some problems that until then had been very difficult to deal with, among which we can highlight the classification of images. The appearance of neural networks, and specifically Convolutional Neural Networks (CNN)\textsuperscript{[1–3]}, introduced a significant improvement in this task. Throughout this article, we will try to show a possible application of current quantum computers in such classification tasks by creating a hybrid model and classical convolutional networks that have already demonstrated their potential in this field.

The emerging field of hybrid quantum-classical algorithms joins CPUs and QPUs\textsuperscript{[4]} to speed up specific calculations within a classical algorithm. This allows for shorter quantum executions that are less susceptible to the cumulative effects of noise and run well on today’s devices. This is why we intend to explore the performance of a hybrid convolutional neural network model that incorporates a trainable quantum layer by one hand, and by the other, effectively replacing a convolutional filter in both quantum simulators and QPU.

We propose to design a trainable quantum convolutional filter in a hybrid neural network, appealing for the NISQ era, inspired by these papers \textsuperscript{[5, 6]}, but generalizing these previous works, and using cloud-based QPU.

II. RELATED WORK

Since Alan Turing demonstrated in 1936 that there exist non-computable problems, the interest in creating new ways of solving them has grown remarkably. This, together with the consequences of the well-known Moore’s Law, gave way to the idea of building quantum computers. Throughout these last decades, the superiority of these new computers has been demonstrated to solve some specific problems such as the factorization of prime numbers through Shor’s algorithm\textsuperscript{[7]} or the search in disordered sets with Grover’s algorithm\textsuperscript{[8]}, although all this limited to the number of qubits available. We are currently in the NISQ\textsuperscript{[9]} era in which we have computers between 50-100 qubits, opening the way to the emerging field of hybrid quantum-classical computing. Within this, different algorithms have been developed, such as VQE\textsuperscript{[10]}, QAOA\textsuperscript{[11]} or, in which we will focus, Quantum Machine Learning (QML)\textsuperscript{[12–17]}.

Since the emergence of deep learning, CNN has helped accelerate image processing, NLP, or even chemistry; the use of CNN has flooded almost every industry. The Ref.\textsuperscript{[18]} proposes using continuous filter convolutional layers to model local correlations without requiring that the data be in a grid. They applied a new deep learning architecture that models quantum interactions in molecules. The following Ref.\textsuperscript{[19]} verifies whether the QCNN model can efficiently learn compared to CNN through training using the MNIST dataset through the TensorFlow Quantum platform. While Ref.\textsuperscript{[20]} provides a generic framework for simultaneously encoding and decoding procedures. The said framework helps to find that the significant schema surpasses known quantum codes of comparable complexity.

In Ref.\textsuperscript{[5]} the author presents a new type of transformational layer called quantum convolution that operates on the input data by locally transforming the data using a series of random quantum circuits, in a similar way to the transformations performed by layers of Random convolutional filters. Although his work showed that QNN models had higher test set accuracy and faster training
compared to purely classical CNNs, it is worth saying
that this approach uses the fixed quantum filter, not vari-
ational. And what it does is pre-process the data before
applying it to the network.

After analyzing the state of the art in CNN deeply, we
did not find any approach like the one we propose. Create
a variational quantum filter (QFilter), taking advantage
of all the classical CNN’s experience and only substitute
the scalar product for a quantum one. What we intend
with this proof of concept is to make sense of the hy-
brid computing platform and align with the Pennylane
philosophy- Training a quantum computer the same way
as a neural network.

III. CNN

Convolutional neural networks [2, 3] are used mainly to
treat problems with a large number of images character-
istic of Deep Learning. Regarding its operation, it could
be divided into two stages. The first one will be in charge
of passing the image through some filters to create new
images that facilitate understanding the network. After
this, we will give the previous phase’s output through a
full-connected neural network capable of learning thanks
to a cost function. Finally, it is worth mentioning that
both the layers of the neural network and the filters are
parameterized, so it is expected that, with the training
process, they will be updated towards the desired values.

A. Convolutional filters

Focusing on the first of the phases described, we will
talk about convolutional filters. The process followed to
transform the image with these follows a simple process.
Initially, a filter is defined as an $n \times n$ matrix, then a
window of the same dimension will go through the image
performing the operation shown in Fig. (1).

\[
I : \begin{bmatrix} I_1 & I_2 & \cdots & I_n \end{bmatrix}
\]

\[
F : \begin{bmatrix} F_1 & F_2 & \cdots & F_n \end{bmatrix}
\]

\[
S = \sum_{i=1}^{n} I_i \cdot F_i
\]

Figure 1. Suppose we take $n = 3$, that is, the filter will have a
total of 9 elements. The result after performing the operation
is to obtain $\sum_{i=1}^{9} I_i F_i$.

As the filter is moved, the solution image’s size will be
reduced compared to the initial input. There are padding
techniques with which it could be possible to make the
final image retain its size, but we have chosen not to carry
out this process to reduce the number of parameters. In
this way, given an image of dimension $l \times w$ and a filter
of $n \times n$, we will obtain an output of $\left[\frac{l}{n}\right] \times \left[\frac{w}{n}\right]$.

IV. OUR MODEL

In this article, we will create a convolutional network to
classify the MNIST dataset [21] (set of images with hand-
written digits 0 to 9) and the fashion MNIST dataset [22].
We will work with a convolutional layer in which we will
apply 4 filters, and later we will connect it to a neu-
ral layer whose output will have dimension 10, one for
each digit. The outcome represents each class’s proba-
bility, and we will say that an image belongs to the class
whose probability is more outstanding. In these cases in
which we want to approximate a probability, the crossed
entropy is applied to calculate the defined error $E$ as fol-

\[
E = \frac{-1}{M} \sum_{i=1}^{M} y_i \log(\hat{y}_i)
\] (1)

Where $M$ is the number of classes, $\hat{y}_i$ is the probability
obtained from the class $i$, and $y_i = 1$ if the label is $i$
or 0 otherwise. Up to this point, the process followed
could be interpreted as a classical development applying
convolutional networks. However, we have decided to
carry out a quantum approximation, replacing, in this
case, the classical filters with quantum procedures. The
birth of this idea arises when representing the elements
of the image and the filter as vectors:

\[
I = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ F_2 \\ \cdots \\ F_n \end{bmatrix}
\] (2)

In this way, it is easy to realize that the operation
carried out is nothing more than the scalar product of
said vectors. Therefore, we could denote it according to
Dirac’s notation as $\langle I | F \rangle$. Looking at it this way, that
starts to get a sense of the idea behind filter quantization.

A. Quantum filters

Let us define a quantum filter as that filter that ex-
ecutes the dot product $\langle I | F \rangle$ in a quantum circuit. To
transfer this process to the circuit, we must first encode
input \( I \). By definition, all its elements are numbers between 0 and 1 (grayscale), so the most logical embedding is to encode said values at angles with gate \( R_Y \). In this way, the filter size will determine the number of qubits in the circuit, requiring \( n^2 \) qubits. We can then represent the input as \( R_Y(I) |0\rangle^n \). Regarding the filter, we design the ansatz determined by \( n^2 \) parameters shown in Fig. (2).

![Figure 2](image)

Figure 2. Schemes of the ansatz that we use to achieve the quantum convolutional adaptive filter. We use two one-qubit gates \( RX, RY \) and a \( CNOT \) gate to accomplish the entanglement.

For simplicity, we will denote this ansatz as \( F(\theta) |0\rangle^n \). Having all this notation, what we want is to obtain the dot product, that is \( \langle 0^n | I^1 F(\theta) |0^n \rangle \).

To calculate this product, we are going to make two different approximations. In the first way, we actually compute \( |\langle 0^n | I^1 F(\theta) |0^n \rangle|^2 \) and consists of constructing the circuit \( |I^1 F(\theta) |0^n \rangle \) and obtaining the probability of measuring \( |0\rangle^n \). In this first scenario, it was decided to calculate the global probability that \( |0\rangle^n \) would appear over the rest of the possibilities. Still, it is equivalent (although more efficient in several shots) to calculate the individual probabilities that each qubit takes the value \( |0\rangle \) and then multiply all the results obtained.

However, although this form may be valid, we will never obtain negative values when calculating the expression’s squared modulus when, in fact, the scalar product could have been. For this, we will carry out a second approximation to calculate \( \langle 0^n | I^1 F(\theta) |0^n \rangle \). In this new scenario, we will calculate the exact value of the inner product (a complex number). However, even if we have the exact value, this does not matter since the neural network will work with real numbers. In this case, what we will do is calculate the real part of said value. For this, we will use the Hadamard Test (see Fig. (3)) based on building the following circuit.

![Figure 3](image)

Figure 3. We create a random variable with this circuit whose expected value is the expected real part of a quantum state’s observed value concerning \( \text{Re} (\langle 0^n | I^1 F(\theta) |0^n \rangle) \)

by equation (3).

\[
\text{Re} (\langle 0^n | I^1 F(\theta) |0^n \rangle) = 2P(0) - 1 \tag{3}
\]

Where \( P(0) \) is the probability of obtaining 0 when measuring in the previous circuit.

**B. Implementation**

To carry out the project, we have taken advantage of the fact that Pennylane [23] provides an interface for Tensorflow and, in this way, use functions already created efficiently. Thanks to this, we have built a hybrid training flow in which we use two different optimizers: SGD [24, 25] to update the classical parameters and Adadelta [26] for the quantum ones. We have made this distinction since, during the experimentation, we observed that the gradient of the quantum parameters was imposed on the classics without letting that part learn; however, we guarantee double training during the experimentation.

**V. RESULTS**

Before comparing our QFilter in a general way with the classical performance and comparing it with other contributions seen, it is of the utmost importance that we validate its operation globally and affirm that QFilter does meet our expectations and works as we expected. We wanted to analyze the entire test set based on our training accurately through the test benches we made. Once the model is built, the first step is to see that our model is capable of learning and generalizing. To do this, we take 50 images from our dataset (MNIST) and trained it for 30 and 50 epochs obtaining the following results Fig. (5).

In this case, starting at the 15th epoch, the model begins to have a precision of 40%. It may seem like a bad result on the surface, but let us analyze it. As we said before, we have used 50 images, and we are trying to divide between a set of 10 different classes; that is; we are using no more than five images for each class. Despite this, we have run this circuit with a traditional model reaching similar limits as a check.
On the other hand, when implementing the model for the first time, with the filters’ size $2 \times 2$, we faced difficulty with the number of operations (dot products). To understand this, let us calculate said number of operations ($S$) in a generic way for $p$ image of $l \times l$ and $m$ filters of $n \times n$ during $t$ epochs.

$$S = \left(\frac{l}{n}\right)^2mnt$$

Therefore, in our initial case of 30 times 4 filters of size 2 and with 50 images of $28 \times 28$, we performed a total of 1,176,000 operations. We decided to increase the filter size to 4 to solve this problem, reducing 294,000. The reduction in the number of operations is significant, and really what we are doing is transferring the computational load of the classical part to the quantum scalar product because by increasing the size of the filter, we raise the size of the circuit from 4 qubits to 16. We were looking for this since we can easily enhance this gradient calculation with the parallelism offered by Amazon Braket with Pennylane API\[27\].

![Diagram](image)

Figure 4. With Amazon Braket we can enhance the quantum scalar product that we had defined as $\langle 0^n | I \, T \, F(\theta) | 0^n \rangle$ in parallel. With the API, Pennylane-Braket we can use the high-performance state vector simulator (SV1)\[28\] that is designed with parallel execution to run in parallel all the circuits needed to compute a gradient.

After experimenting with the different test benches, we can present the results. Figure 5 shows the good behaviour of the variational filter in front of the classical one.

Figure 5 to Fig. 7 show the QFilter results as quantum variational with 50, 60 and 100 epochs.

**VI. DISCUSSION**

Apart from the results, we want to highlight the points that we consider of the utmost importance in this proof of concept. It is worth remembering that one of the goals of this proof of concept is to offer an approach to using hybrid convolutional filters by harnessing the full potential of current convolutional networks and embedding the quantum part only by changing the scalar product. Based on the experiments we did, we found it essential
to define the following guidelines to achieve the code’s adequate scalability.

1. Spin up larger jupyter notebook instances, for example the type ml.c5.2xlarge (8 vCPU, 16 GiB Memory) to speed up CPU optimization and quantum simulation time.

2. Compare the computation time of remote/local simulator.

3. Increase the number of qubits (filters of size 4x4 and 6x6).

4. Batch parallelization of the quantum circuits at the gradient and convolution translation operation levels.

By increasing the filter window size, we are simulating more qubits, therefore simulation time also increases exponentially with the number of qubits, at least in full wave-function or state vectors simulators. The gradient computation involves many quantum circuits executions, it scales approximately as $2 \times m$ where $m$ is the number of trainable parameters, due to the parameter shift rule used to calculate the trainable quantum filter gradients that propagate through the network, and that really blows up the running time, although we expect high performing remote simulators that are able to batch/parallelize quantum circuits to help. Here is a benchmarking for a fixed set of hyper-parameters of the 16-qubits quantum filter, to get a sense of the running time. Another aspect to consider is network latency between the CPU and remote QPU (high performance simulators), this can quickly be the major overhead as the feedback loop between classical and quantum computation is iterated many, circuit device executions that are sent to cloud based instances of QPU and simulators must be minimized. For that, the geographical localization of both processors is a must.

The code Ref.[29] has been prepared to work for large filter size, so the potential of the API (PennyLane with Amazon Braket) can be exploited in order to parallelize the execution of the Qfilter.

A. Benchmark

To test our algorithm, we decided to compare with a similar case already studied Ref.[6]. In this case, the MNIST is also used, and a fixed quantum filter is applied; that is, it does not train any parameter. The reason for taking random parameters is defined in Ref.[6]. It is detailed that this type of filters is suitable for detecting vertices, particularly in this dataset; this becomes a relevant skill. As we can observe Fig.[9], both models end up converging around a 40% precision, which, as we have said before, given the small number of images with which we are training, is a successful result. We can also observe in detail the comparison between the variational, fixed and classical filter. It can be seen the best behaviour of our variational filter.
Figure 9. The blue graph marks the precision of our model throughout the training process, the red graph refers to the learning of the quantum fixed model, while the green one is the classical scalar product.

Also, to attest to the proper functioning of the QFilter, we test it with the MNIST fashion dataset Fig. (10). The results have been satisfactory considering the characteristics of said database. We also did the tests with this same dataset, but with the filter fixed to observe its behaviour, and we noticed the variational filter's importance. Since with the fixed filter, it is possible to detect the contours acceptably, in the case of clothing images, as is the point of fashion MNIST, its limitation is seen.

| QFilters | MacBookPro | Amazon Braket |
|----------|------------|--------------|
|          | 8-Core Intel | ml.t3.medium | ml.c5.2xlarge |
| 2x2      | 3.52        | 19.30        | 4.52          |
| 4x4      | 140.05      | 229.30       | 170.52        |
| 5x6      | –           | –            | –             |

Table I. Table that highlights the importance of latency, the type of instance on which we run our project. This table summarizes in some way the simulation and comparison strategy to take into account for future benchmarks. In addition to all this, the device to be used must be taken into account. All the values are in second and are the processing image time for the variational filter.

One aspect worth finding out is that the accuracy is better for the quantum filter even when the cost function is not. This is an essential advantage because it could mean that hard training are not necessary to achieve better results.

We must emphasize that, due to the era in which we are (NISQ) and due to the novel quantum architecture, the variational circuit’s simulation is much more expensive than the classical one. Table (I) shows us the processing time in seconds of each image, depending on the filter and the device of the quantum algorithm’s execution.

**B. Conclusion and future directions**

Throughout this proof of concept, we have been able to test different strategies to tackle somewhat larger than usual hybrid programming problems. Although the number of operations grows, the results obtained have been satisfactory given the experiments carried out, keeping in mind the small number of images used for training. For this reason, we consider that this procedure would be competent in situations of great uncertainty. Another way to continue exploring is the parallelization of the classical part in order to reduce time and study the behaviour in larger images.
C. acknowledgements

We want to thank Yoshitaka Haribara, Cedric Lin, Pablo Nicolás Nuñez Pölcher and Ricardo García for the support and discussions during this proof-of-concept development. We would also like to thank Xanadu/Pennylane and its team for hosting Qhack, for all the effort put into this super event, and all the sponsors involved. Especially Amazon Web Services, to provide access to its resources (S3, Amazon Braket) to carry out part of this project.

[1] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET). IEEE, August 2017. doi:10.1109/icengtechnol.2017.8308186

[2] Keiron O’Shea and Ryan Nash. An introduction to convolutional neural networks, 2015. arXiv:1511.08458

[3] Junhua Liu, Kwan Hui Lim, Kristin L Wood, Wei Huang, Chu Guo, and He-Liang Huang. Hybrid quantum-classical convolutional neural networks. arXiv preprint arXiv:1911.02998, 2019.

[4] Parfait Atchade-Adelomou, Elies Golobardes-Ribe, and Xavier Vilasis-Cardona. Using the parameterized quantum circuit combined with variational-quantum-eigensolver (vqe) to create an intelligent social workers’ schedule problem solver, 2020. arXiv:2010.05863

[5] Parfait Atchade-Adelomou, Daniel Casado-Fauli, Elies Golobardes-Ribe, and Xavier Vilasis-Cardona. Quantum case-based reasoning (qcbr), 2021. arXiv:2104.00409

[6] Parfait Atchade-Adelomou, Alba Cervera-Lierta, Elies Gil, and Johan van den Brink. A quantum-classical convolutional approach for image classification. In 2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, October 2020. doi:10.1109/ictc49870.2020.9289439

[7] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. Nature Physics, 15(12):1273–1278, Aug 2019. URL: http://dx.doi.org/10.1038/s41567-019-0648-8 doi:10.1038/s41567-019-0648-8

[8] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre van Schaik. EMNIST: Extending MNIST to handwritten letters. In 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, May 2017. doi:10.1109/ijcnn.2017.7966217

[9] Minmin Liu, Bu-Jia Cao, and Guangming Gu. A quantum approximate optimization algorithm, 2014. arXiv:1411.4028

[10] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. An introduction to quantum machine learning. Contemporary Physics, 56(2):172–185, Oct 2014. URL: http://dx.doi.org/10.1080/00107514.2014.964942 doi:10.1080/00107514.2014.964942

[11] Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. Physical review letters, 122(4):040504, 2019.

[12] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. Nature, 549(7671):195–202, Sep 2017. URL: http://dx.doi.org/10.1038/nature23474 doi:10.1038/nature23474

[13] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017. arXiv:1708.07747

[14] Maxw, Henderson, Samriddhi Shakya, Shashindra Pradhan, and Tristan Cook. Quantum convolutional neural networks. arXiv preprint arXiv:1911.02998, 2019.

[15] Parfait Atchade-Adelomou, Alba Cervera-Lierta, Elies Gil, and Johan van den Brink. A quantum-classical convolutional approach for image classification. In 2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, October 2020. doi:10.1109/ictc49870.2020.9289439

[16] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. Nature Physics, 15(12):1273–1278, Aug 2019. URL: http://dx.doi.org/10.1038/s41567-019-0648-8 doi:10.1038/s41567-019-0648-8

[17] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre van Schaik. EMNIST: Extending MNIST to handwritten letters. In 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, May 2017. doi:10.1109/ijcnn.2017.7966217

[18] Minmin Liu, Bu-Jia Cao, and Guangming Gu. A quantum approximate optimization algorithm, 2014. arXiv:1411.4028

[19] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. An introduction to quantum machine learning. Contemporary Physics, 56(2):172–185, Oct 2014. URL: http://dx.doi.org/10.1080/00107514.2014.964942 doi:10.1080/00107514.2014.964942

[20] Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. Physical review letters, 122(4):040504, 2019.
[25] Nitish Shirish Keskar and Richard Socher. Improving generalization performance by switching from adam to sgd, 2017. arXiv:1712.07628

[26] Matthew D. Zeiler. Adadelta: An adaptive learning rate method, 2012. arXiv:1212.5701

[27] Amazon Braket PennyLane. Pennylane-braket plugin. https://amazon-braket-pennylane-plugin-python.readthedocs.io/en/latest/ 2021 (accessed February 26, 2021).

[28] Amazon Braket PennyLane. Pennylane-braket plugin. https://docs.aws.amazon.com/braket/latest/developerguide/braket-devices.html 2021 (accessed March 26, 2021).

[29] Guillermo Alonso-Linaje and Parfait Atchade-Adelomou. Qfilter, 2021. URL: https://github.com/

[30] Apple. Macbookpro 15.1. https://support.apple.com/kb/SP776?locale=es_ES 2021 (accessed April 6, 2021).

[31] Amazon Braket. Amazon braket. https://aws.amazon.com/es/ec2/instance-types/ 2021 (accessed February 26, 2021).