ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

Explorations in Quantum Neural Networks with Intermediate Measurements

Lukas Franken and Bogdan Georgiev

*Fraunhofer IAIS - Research Center for ML and ML2R
Schloss Birlinghoven - 53757 Sankt Augustin

Abstract. In this short note we explore a few quantum circuits with the particular goal of basic image recognition. The models we study are inspired by recent progress in Quantum Convolution Neural Networks (QCNN) [12]. We present a few experimental results, where we attempt to learn basic image patterns motivated by scaling down the MNIST dataset.

1 Introduction

The recent demonstration of Quantum Supremacy [1] heralds the advent of the Noisy Intermediate-Scale Quantum (NISQ) [2] technology, where signs of superiority of quantum over classical machines in particular tasks may be expected. However, one should keep in mind the limitations of NISQ-devices when studying and developing quantum-algorithmic solutions - among other things, these include limits on the number of gates and qubits.

At the same time the interaction of quantum computing and machine learning is growing, with a vast amount of literature and new results. To name a few applications, the well-known HHL algorithm [3], quantum phase estimation [5] and inner products speed-up techniques lead to further advances in Support Vector Machines [4] and Principal Component Analysis [6, 7]. Intensive progress and ongoing research has also been made towards quantum analogues of Neural Networks (QNN) [8, 9, 10]. Boltzmann machines use sampling in a connected graph to the examples created by an unknown probability distribution. Here the energy function of an Ising spin system is used. Adiabatic Quantum Computers naturally obey the statistics of such an energy function an hence they are an intuitive option to build on the success of Boltzmann architectures. Their quantum mechanical nature allows AQC to recognize statistical patterns that are classically impossible to catch. On the other hand, hybrids of quantum and classical methods (Variational Circuits) can be used for supervised learning. The feedforward procedure is defined by the configuration of parameters acting on a quantum circuit. These parameters are optimized e.g. by gradient descent where a classical loss function is defined [11].

In this note we study a class of quantum circuits that are inspired by Convolutional Neural Networks (CNNs). Attempts to replicate the CNN design have already been put forward [12]. We note that a well-known issue when trying

*We would like to thank the Fraunhofer IAIS, the Research Center for Machine Learning IAIS, and the Competence Center for Machine Learning Rhine-Ruhr ML2R, for the excellent working conditions. We would also like to thank Prof.C. Bauckhage for the encouragement and fruitful discussions

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

to replicate a neural network with a quantum computer are the non-linearities occuring at every layer. In [12] such issues are addressed by using partial measurement as a tool for reduction of dimensionality and as a way to introduce non-linearities in the system. We explore a related quantum circuit architecture where we view the CNN poolings and non-linearities as appropriate combinations of (unitary) rotations and measurements [1]. Below we study the ability of these models to recognize simple patterns in small-scale image data. The possible advantage of such an approach stems from the number of required parameters, which are for an input of size N of order $\mathcal{O}(\log(N))$.

2 Circuit Exploration and Experiments

CNNs have proven to be effective tools as they are tailored to obtain structural information stemming from local recurrence of values. A CNN usually consists of a sequence of layers that "coarse-grain" the input, while in between fully-connected layers are used for reduction of dimension while preserving the structural information. The quantum-mechanical translation to this procedure we use is inspired by [12].

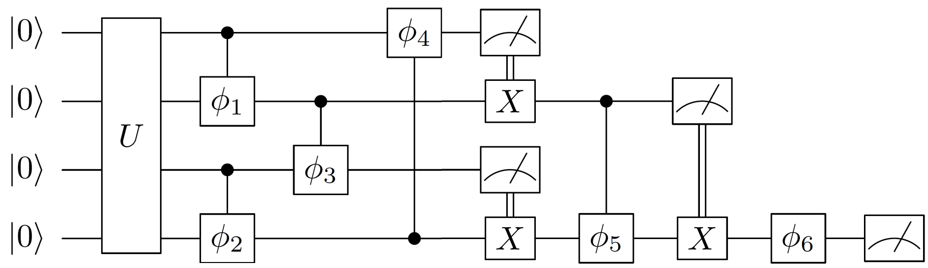

Fig. 1: An overview of the quantum circuit model used in the experiment. After encoding the pattern, controlled rotations are learned. The subsequent measurements reduce the dimension of the problem. The procedure is then repeated.

2.1 The Basic Circuit Model

In our model, four qubits are initiated in the state $|0\rangle$. Then, using the technique proposed in [13], a subsequent unitary operation maps the states to the superposition, which represents our data (see multigate U in figure 1). This operation requires quadratic gates with respect to the input size. Yet, the resulting states are highly precise and for small input sizes, the gate requirements do not exceed the limitations of NISQ devices. After state preparation, we conduct controlled

[1]Some of our experiments are available at: https://github.com/bogeorgiev/quantum-algorithms-explorations

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

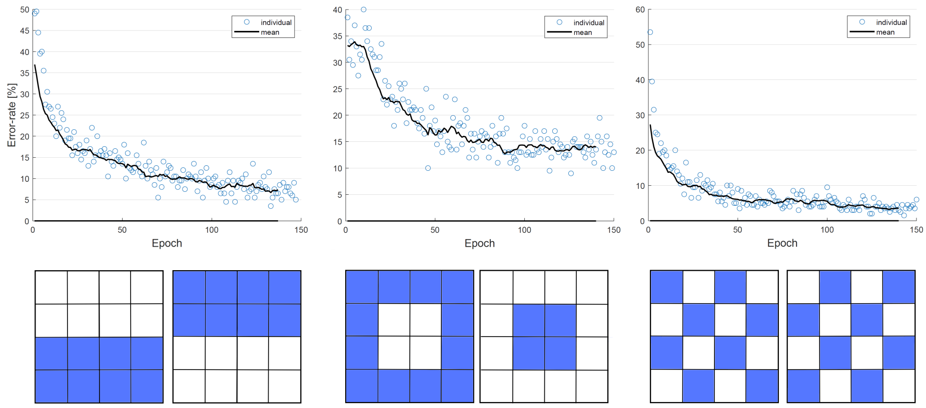

Fig. 2: Bottom: The patterns to be classified. Top: The percentage of miss-classification with respect to the epoch. The dots indicate the resulting error percentage after 200 circuit repetitions. The black lines are the mean of the surrounding points. For all of the patterns quick convergence is observed, where the peak performance of the middle patterns is substantially less precise than for the other patterns.

$R_y(\phi_i)$ rotations given by

$$R_y(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} .$$

Before we conduct the first layer of measurements leading to a reduction of dimension, some structural information should be obtained. We then repeat the procedure on the remaining circuit. Here the nontrivial question arises, what the ideal relation between the entanglement of an input vector and the required control for operations is, as for nonzero entanglement, the pattern might contain information in the non-local correlations between qubits. While product states might be classified best using local operations, highly entangled ones might require multiple controls for gate operations. For a random input vector both are possible, hence in our approach we chose rotations controlled only by one qubit, giving the circuit flexibility for various inputs.

Elaborate toolboxes provide well tuned methods to determine gradients in classical machine learning. In the classical environment, this procedure is particularly easy, because the nonlinear activation functions are designed to have gradients, that are easy to determine. Time evolution according to the Schrodinger equation is governed by unitary operations and hence linear. The description via Schrodinger equation breaks down at the intersection between the quantum and classical world, which can be introduced using measurements. We exploit this to create nonlinearity by projective measurement which is a mapping $|\psi\rangle \mapsto P_m |\psi\rangle / p(m)$, where P_m is a projection to a one-qubit basis state with probability $p(m)$. While providing nonlinearity, this operation still has no

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

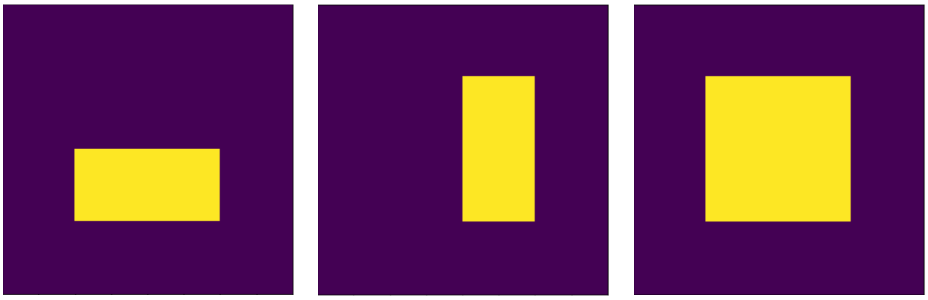

Fig. 3: Means of the down-sampled MNIST set for digits 6, 3 and 8

clear notion of a gradient. As a proxy for a gradient we here use finite differences [11], according to

$$\frac{\partial L}{\partial \theta_j} \approx \frac{L(\theta + \Delta\theta_j) - L(\theta - \Delta\theta_j)}{2\Delta} \tag{1}$$

where θ_j is one of our training parameters and L is the cross entropy given by

$$L(p, q) = -\sum_i p_i \log q_i \ . \tag{2}$$

The cross entropy is evaluated between the softmax function of the circuit's final measurement as obtained after the testing repetitions and the softmax of the actual classification of the example. We evaluate the circuit's quality via the percentage of misclassification.

Figure 2 shows the results of the first 3 experiments for simple pattern recognition tasks. The circuit clearly shows learning for all three tasks with peak performance at 3%, 9% and 1.5%, respectively. This result is in accordance with the intuition on quantum computers being able to deal with large data for small resources, but at the same time delivering a good instead of a perfect solution. Reducing the error to zero would require reaching perfect orthogonality between states before the final measurement. Due to the measurements not being a majority vote as in most classical supervised learning tasks, having nonzero misclassification chance will most probably remain an aspect of minimalist quantum circuits.

2.2 MNIST classification

It has been demonstrated that a quantum circuit can be successfully trained on the down-sampled MNIST data set [8]. Yet, while the optimal error rate of 2% obtained in the cited paper is impressive, large resources are employed to reach such a result. The following table summarizes the resources required compared to the architecture investigated in this note.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

	QCNN ansatz	QNN ansatz [8]
qubits	4	17
trained parameters	5	170

Furthermore, in the outlined MNIST classification experiments [8], a smaller data set was introduced where one works with two digits only and deals with representative samples that are sufficiently distinguishable (w.r.t. some metric). Motivated by this, we cropped the MNIST dataset in the following way. We clustered the (down-sampled to 4×4 pixels) MNIST digits by a k-means procedure and found the corresponding cluster centers - a further normalized picture, for example, of these means is given in Figure 3. In fact, these cluster means were the motivation behind considering the patterns in Figure 2.

Afterwards we selected digit samples that are sufficiently close (in l^2 sense) to these cluster means and thus formed our dataset. Working with the cluster means and digit samples that are very close to the cluster means, the model experiences a similar behaviour to the toy-examples in 2. We expect that with additional fine-tuning and slight extension the model will be able to perform acceptably well even for classification of digit samples that are further away from the cluster means. We study these issues in an upcoming work.

3 Discussion and Further Work

In this short note, we have presented a few explorations that exhibit promising results with a 5-10% error rate, despite circuit architecture being chosen with more concern for the restrictions of NISQ than for the problem setup. We plan to further build upon these initial observations - both in terms of complexity and rigorous investigation. In an upcoming work we intend to analyze further the notion of convolution that was understood in terms of appropriate intermediate measurements. Further theoretical questions arise when one also asks about robustness of gradient descent and "barren-plateau"-related issues [15]. With this in mind, we view the present note as just the beginning step towards further (both theoretical and practical) work in the subject.

References

[1] Arute, Frank and Arya, Kunal and Babbush, Ryan and Bacon, Dave and Bardin, Joseph C. and Barends, Rami and Biswas, Rupak and Boixo, Sergio and Brandao, Fernando G. S. L. and Buell, David A. and et al., Quantum supremacy using a programmable superconducting processor, Nature, 7779, p. 505-510, Springer Science and Business Media LLC, 2019.

[2] John Preskill, Quantum Computing in the NISQ era and beyond, 2018.

[3] Harrow, Aram W. and Hassidim, Avinatan and Lloyd, Seth. Quantum algorithm for linear systems of equations, Physical Review Letters, 00319007, 2009.

[4] Rebentrost, Patrick and Mohseni, Masoud and Lloyd, Seth, Quantum support vector machine for big data classification, Physical Review Letters, 10797114, 2014.

[5] Nielsen, Michael A. and Chuang, Isaac L. Quantum Computation and Quantum Information, 10.1017/cbo9780511976667, 2010.

[6] Schuld, Maria and Sinayskiy, Ilya and Petruccione, Francesco, An introduction to quantum machine learning, Contemporary Physics, 13665812, 2015.

[7] Biamonte, Jacob and Wittek, Peter and Pancotti, Nicola and Rebentrost, Patrick and Wiebe, Nathan and Lloyd, Seth. Quantum machine learning, Nature, 14764687, 2017.

[8] Edward Farhi and Hartmut Neven. Classification with Quantum Neural Networks on Near Term Processors, preprint, arXiv: 1802.06002, 2018.

[9] Dumoulin, Vincent and Goodfellow, Ian J. and Courville, Aaron and Bengio, Yoshua, On the challenges of physical implementations of RBMs, Proceedings of the National Conference on Artificial Intelligence, 2014.

[10] Denil, M and De Freitas, N, Toward the Implementation of a Quantum RBM NIPS Deep Learning and Unsupervised Feature . . . , 2011.

[11] Benedetti, Marcello and Lloyd, Erika and Sack, Stefan and Fiorentini, Mattia, Parameterized quantum circuits as machine learning models, Quantum Science and Technology, 2058-9565, 2019.

[12] Cong, Iris and Choi, Soonwon and Lukin, Mikhail D. Quantum convolutional neural networks, Nature Physics, 1745-2481, 2019.

[13] Lov Grover and Terry Rudolph, Creating superpositions that correspond to efficiently integrable probability distributions, quant-ph/0208112

[14] S. Haykin, editor. *Unsupervised Adaptive Filtering vol.1 : Blind Source Separation*, John Willey ans Sons, New York, 2000.

[15] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush Hartmut Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications volume 9, Article number: 4812 (2018)

[16] N. Delfosse and P. Loubaton, Adaptibe blind separation of sources: A deflation approach, *Signal Processing*, 45:59-83, Elsevier, 1995.

[17] S. Cruces, A. Cichocki and S. Amari, The minimum entropy and cumulants based contrast functions for blind source extraction. In J. Mira and A. Prieto, editors, proceedings of the 6^{th} *international workshop on artificial neural networks* (IWANN 2001), Lecture Notes in Computer Science 2085, pages 786-793, Springer-Verlag, 2001.

[18] F. Vrins, C. Archambeau and M. Verleysen, Towards a local separation performances estimator using common ICA contrast functions? In M. Verleysen, editor, *proceedings of the* 12^{th} *European Symposium on Artificial Neural Networks* (ESANN 2004), d-side pub., pages 211-216, April 28-30, Bruges (Belgium), 2004.

[19] J. V. Stone and J. Porrill, Undercomplete independent component analysis for signal separation and dimension reduction. Technical Report, Psychology Department, Sheffield University, Sheffield, S10 2UR, England, October 1997.