POLYTECHNIC UNIVERSITY OF BUCHAREST
FACULTY OF AUTOMATIC CONTROL AND COMPUTERS
COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

BACHELOR THESIS

A comprehensive study on Integrating Quantum Physics in Deep Learning Models for Optimizations in Computer Vision and Natural Language Processing Fields

QRKT-GAN: Neural Ordinary Differential Equation-Inspired Generative Adversarial Network Model with Numerical Runge-Kutta Methods for Quantum Visual Transformer-Based Generator and Discriminator

Cătălin-Alexandru Rîpanu

Thesis advisor:

Ș.l. dr. ing. Dumitru-Clementin Cercel

BUCHAREST
2024

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Dumitru-Clementin Cercel, for his unwavering support and insightful guidance throughout my research journey. His expertise and encouragement have been essential in exploring the intricate intersection of Artificial Intelligence and Quantum Mechanics, widely referred to in the literature as Quantum Artificial Intelligence. This interdisciplinary field, though initially daunting, is at the cutting edge of technological advancement and innovation in both areas.

I am also profoundly thankful to Prof. Pantelimon-George Popescu for his advice, recommendations, and exceptional lectures on Quantum Computing. His mentorship has been instrumental in building a robust understanding of this groundbreaking paradigm, which promises new solutions to problems previously deemed insurmountable by classical methods.

Furthermore, I wish to acknowledge the significant contributions of the academic teaching assistants and collaborators encountered during my bachelor's studies. Their guidance during challenging times has been crucial to my academic development. Last but not least, I am also grateful for the computational resources provided by the Computer Science and Engineering Department, which were essential for the design and implementation of this project.

In addition, I extend my appreciation to my friends for their continuous support, ideas, and advice, especially during moments when inspiration was scarce. Lastly, my heartfelt thanks go to my parents for their enduring sacrifices and unwavering support throughout this journey into the unknown. Their belief in me has been a constant source of motivation.

ABSTRACT

Deep Learning models, such as Generative Adversarial Networks (GANs) [1] and Visual Transformers (ViTs) [2], have demonstrated remarkable results across various domains in Machine Learning and Artificial Intelligence, including Object Classification, Image Segmentation, Sentiment Analysis, and Machine Translation. These models are pivotal in advancing systems that require high precision and a deep understanding of complex data across a variety of tasks in both Natural Language Processing (NLP) and Computer Vision (CV).

However, the effectiveness of Deep Learning models comes with significant challenges: they require an extensive number of parameters to learn and extract meaningful features from real-world data. Additionally, these models need vast amounts of information to achieve desired performance levels. Obtaining such large sets can be difficult as real-world data is often not publicly available and can be challenging to collect and curate. This results in substantial computational resource requirements for both training and hyperparameter optimization, often achieved through exhaustive techniques such as grid search.

To address these challenges, this thesis proposes a novel hybrid Generative Adversarial Network architecture that employs Quantum Visual Transformers (QViTs) as both the Generator and Discriminator. Visual Transformers are selected for their superior ability to manage intricate data representations. A key innovation in this architecture is the integration of Ordinary Differential Equation (ODE) [3] solvers as Encoders, enhancing the model's capability to capture temporal dynamics and complex data structures, and improving the residual connections within the transformer architecture to mitigate the vanishing gradients problem even more.

Moreover, this architecture incorporates Variational Quantum Circuits [4] within both the Self-Attention Mechanisms [5] and the Multi-Layer Perceptrons (MLPs) [6] of the Visual Transformers. By leveraging the principles of Quantum Mechanics, these Quantum circuits can perform complex algebraic operations more efficiently than classical methods, offering a significant computational advantage.

The performance of this hybrid model is benchmarked against SoTA purely classical baselines from the literature using datasets from both CV and NLP areas. Specifically, the model is tested on CIFAR-10 [7], CIFAR-100 [7], MNIST [8], IMDb Reviews [9] and ILSVRC 2012 [10] datasets. The Quantum model is successfully trained and tested through numerical simulations. The results indicate that this hybrid approach achieves comparable classification and generation performance to the classical baseline, while requiring fewer trainable parameters.

Furthermore, the reduced parameter count in this hybrid model opens up the possibility of running it on real Quantum hardware for both training and inference. This feasibility is a significant breakthrough, as it implies that Quantum-enhanced models can be trained and deployed on actual Quantum computers, which are currently limited in terms of the number of qubits and operational fidelity.

This thesis demonstrates the potential of integrating Quantum Computing, especially Quantum mechanics, with advanced Deep Learning architectures to create more efficient and powerful networks which can significantly reduce computational costs while maintaining high performance, paving the way for more scalable and effective AI applications.

CONTENTS

1 Introduction

1.1 Context

Artificial Intelligence models, particularly Deep Learning ones, have made significant contributions to solving real-world tasks, greatly improving human lives in various fields, such as Medical Image Recognition [11]. Despite their impressive capabilities, Deep Learning models come with substantial drawbacks regarding computational resources and effort. Achieving high performance with these models necessitates learning millions to billions of parameters, also called weights or artificial neurons, which demands considerable resources and preparation time. This limitation also has negative environmental impacts due to high power consumption.

Over the years, researchers have developed numerous solutions to mitigate the problem of minimizing the number of parameters using interesting classical algorithms and techniques. These include, for example, specialized activation functions for neural layers like Rectified Linear Unit [12], Leaky ReLU [13], Gaussian Error Linear Unit [14], Softmax [15], Sigmoid [16] and Hyperbolic Tangent [17]. Additionally, effective optimizers such as Adam [18] and Stochastic Gradient Descent [19] have been utilized to combat this limitation, along with methods like Learning Rate Schedulers [20], Weight Decay [21], and Dropout [22] to reduce also the chances of model overfitting. Over the years, various architectural designs and network combinations have emerged to address these challenges. However, the tradeoff between high computational resource consumption and performance remains a difficult issue, especially when scaling up the dimensions of available noisy real-world datasets used for training. It is unlikely[1] to be efficiently resolved in the near future using solely classical approaches.

A segment of Computer Science researchers, in collaboration with physicists, have taken a bold and innovative approach to these challenges by exploring solutions from a physical perspective, leading to the emergence of Quantum Computation and Quantum Information [23]. Leveraging the principles of Quantum mechanics, AI research has begun to explore Deep Learning optimizations using Quantum elements, such as Variational Quantum Circuits [4], to develop hybrid Deep Learning models. By combining classical and Quantum methodologies, new experiments can be conducted to address performance and resource demands more efficiently, thus utilizing the unique capabilities of Quantum hardware. This involves training these models on it, also known as Parameterized Quantum Circuits [24], where the parameters are referred to as the qurons of the Quantum model.

Quantum Computing leverages properties like superposition and entanglement, enabling the execution of complex algebraic operations that are infeasible for classical computers. Quantum algorithms, such as Grover's Algorithm [25], Quantum Phase Estimation [26], Quantum Fourier Transform [27] and Deutsch-Jozsa Quantum Parallelism [28] provide exponential speedups for certain tasks. Integrating these Quantum principles with Deep Learning can significantly reduce the number of parameters and computational resources required while maintaining or even enhancing model performance. This interdisciplinary approach, also known as Quantum Artificial Intelligence, opens up exciting possibilities for the future of AI, promising advancements that could transform how complex problems are approached and solved in ways that classical computers, with their current architecture, could never achieve.

[1] https://www.theverge.com/24066646/ai-electricity-energy-watts-generative-consumption

1.2 Problem

Over the years, numerous algorithms, techniques, network architectures, and methodologies have been proposed in the field of Deep Learning to tackle a wide array of tasks, such as Object Classification, Synthetic Data Generation and Detection, across various domains including Natural Language Processing, Computer Vision, Domain Adaptation, and Knowledge Distillation. These approaches have achieved high performance metrics like Accuracy, Area Under the Curve (AUC), Receiver Operating Characteristic (ROC), and F1 scores. The advent of High Performance Computing (HPC) facilities, particularly classical parallelism, has enabled Deep Learning models to process billions of examples from noisy real-world datasets effectively, thus yielding impressive results.

Despite these advancements, a significant drawback persists: the enormous number of neurons and the corresponding computational resources required by these models. Traditional solutions in classical Deep Learning, such as Cross Validation [29], Hyperparameter Tuning techniques like Grid Search [30], Random Search [30], or Keras Tuner [31], and Layer Augmentation until model performance plateaus or overfits, are commonly employed. To mitigate overfitting, regularization techniques such as Ridge [32], Lasso [32], Elastic-Net [32] or Dropout [22] can be applied. However, these methods often involve exhaustive hyperparameter searches, which are extremely time-consuming, especially when dealing with validation sets comprising billions of data points for complex tasks. Additionally, one can apply group sparsity regularizers on network parameters, where each group acts on a single neuron, thus reducing the number of parameters by up to 80% [33] while retaining or even improving the network accuracy.

Another method employed to address these challenges is Transfer Learning [34], which involves finding and loading a pretrained model as a starting point and then fine-tuning it on a specific dataset. This approach can help achieve good performance with less effort on architecture design and expedite the training process. However, the availability of suitable pretrained models is limited, and fine-tuning may yield only decent, if not disappointing, results, due to differences in the data distributions between the pretraining and fine-tuning datasets.

The primary challenge lies in balancing time and power consumption on classical systems to make robust and valid neural architectural choices. Furthermore, reducing the number of neurons too drastically is not a solution as it can lead to underfitting and degraded performance, making it difficult to solve complex and challenging tasks efficiently. The iterative process of tuning hyperparameters and optimizing model architectures is computationally intensive and often results in a significant trade-off between resource consumption and model performance.

Moreover, the reliance on vast amounts of training data for achieving desired performance further exacerbates these issues. Real-world datasets are often difficult to obtain and may not be publicly available, which complicates the development and fine-tuning of effective Deep Learning models. The extensive time and computational costs associated with hyperparameter optimization and model training present significant obstacles to advancing AI capabilities while maintaining efficiency and sustainability. This trade-off between computational resource consumption and performance remains a persistent and challenging problem in the field of Deep Learning, highlighting the need for innovative solutions that can overcome these limitations.

1.3 Objective

In the current landscape of Deep Learning research, there is an ongoing investigation into whether integrating models into a purely Quantum environment can significantly reduce power consumption and the number of trainable parameters without sacrificing performance. This exploration leverages techniques such as the Ansatz [35] and Variational Quantum Circuits [4]. As classical supercomputers continue to increase in power to accommodate newer, more complex deep models, their power consumption scales almost exponentially. Conversely, while the computational power of Quantum machines scales exponentially, their power consumption scales linearly. This interesting information presents a compelling case for Quantum Computing: once Quantum computers achieve fault tolerance, qubits could be efficiently used as artificial neurons in neural networks. Quantum-inspired techniques could allow networks to operate with a vast number of neurons per layer at minimal energy cost, thereby significantly reducing overall energy consumption[2].

Building on the challenges outlined in the above section, the primary goal of this thesis is to explore Deep Learning models, both pure Quantum and hybrid, that leverage neural ODE-based architectures [36] inspired by non-trivial Variational Quantum Circuits [4] within the context of Visual Transformers [2] and their complex Encoders [37] and Decoders [38]. The most important objective is to minimize the number of parameters used by deep network layers utilizing the idea of replacing classical linear projection layers in Multi-Head Attention [5] subroutines and Multi-Layer Perceptrons [6] with Quantum Circuits. Additionally, low truncation error Neural Ordinary Differential Equation [36][39] techniques will be employed to further optimize Transformer Layers, namely Runge-Kutta Methods [40] of 1st, 2nd, 3rd, and 4th order. An analysis of Runge-Kutta 4th order method optimization which is not theoretically grounded in standard RK methods, but in terms of learning, will also be included [39].

To better understand the rationale and motivation behind the proposed Deep Learning model in this paper, it is necessary to compare these Quantum ODE-based Transformers configurations with their classical counterparts using datasets mentioned in the Abstract. The model utilizing the optimized Runge-Kutta 4th order method will be leveraged within the context of Vision Transformers (ViTs) [2] and integrated into a Generative Adversarial Network (GAN) [1] architecture. In this configuration, both the Generator and Discriminator will be Quantum Visual Transformers[3], incorporating the optimized Neural ODE solver to generate realistic and informative synthetic data without being limited to specific types of distributions. This will be achieved using techniques such as Data Augmentation [41] and Image Recognition at Scale [42] in the context of ViTs [2].

The final proposed architecture in this thesis, named QRKT-GAN, will be tested on a real-world consistent dataset, namely CIFAR-10 [7], and compared using the same data with a renowned classical counterpart, TransGAN [43], which is a strong architecture built entirely free of Convolutions. This comparative analysis will provide another argument for the possible advantages of integrating Quantum Computing with advanced Deep Learning neural networks under the right context.

[2]https://www.eetimes.eu/how-Quantum-Computing-can-help-make-ai-greener/

[3]https://openreview.net/pdf?id=p7xPXoKB0H

1.4 Paper Structure

The next chapter delves into the most relevant and important classical and Quantum algorithms and techniques developed over the years to leverage both Quantum and hybrid logic in Deep Learning for model optimization. This chapter will illustrate how these both hybrid and pure Quantum neural networks are designed to be compatible with real Quantum hardware provided by specialized vendors such as IBM [44], Google [45], Microsoft [46], and Amazon [47]. It will explore the advancements that make it possible to run complex Deep Learning models on Quantum computers, highlighting the interplay between Quantum Computing and Deep Learning for neural depth decreasing while not losing too much performance.

Section three will discuss the foundational concepts of Artificial Intelligence, Machine Learning, and Deep Learning. It will also provide a comprehensive overview of Quantum Computation and Quantum Information [23], and will also introduce the emerging field of Quantum Artificial Intelligence [48]. This section aims to explain how Quantum principles can enhance the capabilities of AI solutions in addressing high-demanding tasks, some of which are classified as NP-Hard [49] or NP-Complete [50]. By understanding these fundamentals, the reader will gain insight into the potential synergies between Classical and Quantum Computing, which together form a duality in the world of Artificial Intelligence.

Section four will present an in-depth analysis of the proposed Ordinary Differential Equation-based Quantum neural networks. This includes their motivation, underlying concepts, Mathematical and Quantum frameworks, physical limitations, and future potential within their problem domain environment. The chapter will also discuss the new ODE-based GAN architecture in detail, explaining how it incorporates the traversed knowledge to minimize the solution of the optimization problem outlined in the Context subsection, while maintaining acceptable performance in Synthetic Data Generation.

The fifth chapter will detail the methodologies adopted for conducting the experiments. This includes a comprehensive description of the datasets used, their internal characteristics, and the reasons for their selection. It will compare various Quantum ODE-based configurations with their classical counterparts to fully understand the benefits and limitations of Quantum-enhanced models. Additionally, the chapter will provide a detailed comparison between the proposed QRKT-GAN model and its classical baseline, TransGAN [43], using the CIFAR-10 [7] dataset. This analysis will highlight the benefits and challenges of integrating Quantum Computing into Deep Learning, presenting it as a viable solution for specific optimization problems.

The final chapter will summarize the main ideas presented in the thesis, highlighting the relevance of the results and drawing key conclusions. It will also propose future work and outline potential directions for further research in the field of Quantum Deep Learning models, focusing on power consumption and performance optimization. This chapter will emphasize the importance of continued exploration in combining Quantum Computing with Machine Learning to achieve more robust, efficient and powerful Artificial Intelligence solutions.

2 Related Work

The relentless demand for improving human lives continues unabated, with increasingly complex and data-intensive needs emerging. This necessitates the development of sophisticated models capable of delivering accurate and reliable results. However, the complexity and depth of these models often result in significantly high computational resource consumption, which has environmental implications due to the energy required for such intensive processing.

Over the years, numerous classical, Quantum, and hybrid methods have been proposed to address the optimization problem of minimizing model depth without significantly sacrificing performance. One notable approach is real-time learning of the number of neurons in deep networks using structured sparsity during training, which dynamically adjusts the network structure to optimize performance without excessive resource use [33]. Another method involves crafting specialized architectures that maintain a constant computational budget while slightly increasing depth to improve performance. This is achieved by dimensionality reduction before applying expensive Convolutions [20] with larger patch sizes [51]. Unfortunately, Convolutions [20] have been demonstrated [52] to be less effective in capturing global feature interactions, focusing primarily on local patterns in Image Processing, compared to the Self-Attention [5] mechanism utilized in Transformers [2].

In addition, some techniques focus on minimizing weights during training by splitting network weights into sets or hierarchies of multiple groups, each using disjoint sets of features. This allows parallelization, even in an embarrassingly manner, as each subnetwork operates independently, enhancing computational efficiency [53]. Alternatively, population-based heuristic algorithms, such as Particle Swarm Optimization (PSO), have shown promise in optimally determining the number of parameters without exhaustive searches like grid search [30], and thus saves valuable computational resources during the tuning process of these Deep Learning models [54]. Another approach involves reducing model parameters in deep neural networks via product-of-sums matrix decompositions, which decompose linear operators as a product of sums of simpler linear operators [55]. For Deep CNN models, kernel-sharing between multiple convolutional layers can also be applied. Kernel-sharing is possible only between "isomorphic" layers, i.e., layers having the same kernel size, input, and output channels [56].

Despite their advantages, these classical methods have inherent drawbacks. Designing neural architectures from scratch to meet specific requirements is challenging and time-consuming. Moreover, methods like Particle Swarm Optimization [54] require optimization of additional parameters, such as activation functions [16] [12] [13] and the number of epochs, to be truly effective. Quantum Neural Networks (QNNs) [57] and Quantum Algorithms [4] offer new avenues for addressing these limitations, providing potential solutions for power-efficient Deep Learning. Various Quantum frameworks, technologies, GPU-based numerical simulators, and access to real hardware have been developed to facilitate the implementation of Quantum phenomena in Machine Learning.

However, the Quantum approach shares a common challenge with classical methods: the need for specific adaptations for each type of Deep Learning model. In the classical domain, the entire architectural framework must be meticulously analyzed to prevent performance degradation. In Quantum domain, constructing valid Variational Quantum Circuits (VQCs) [24] that efficiently mimic classical model functions is crucial to avoid the same pitfalls.

3 Background

4 Proposed Solution

5 Experiments

6 Conclusions and Future Work

7 Bibliography

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[3] Zipeng Fan, Jing Zhang, Peng Zhang, Qianxi Lin, and Hui Gao. Quantum-inspired neural network with runge-kutta method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17977–17984, 2024.

[4] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

[5] Elena Voita, David Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. *arXiv preprint arXiv:1905.09418*, 2019.

[6] Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588, 2009.

[7] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

[8] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2, 2010.

[9] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[10] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[11] Yun He, Ziwei Zhu, Yin Zhang, Qin Chen, and James Caverlee. Infusing disease knowledge into bert for health question answering, medical inference and disease name recognition. *arXiv preprint arXiv:2010.03746*, 2020.

[12] Yuanzhi Li and Yang Yuan. Convergence analysis of two-layer neural networks with relu activation. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[13] Jin Xu, Zishan Li, Bowen Du, Miaomiao Zhang, and Jing Liu. Reluplex made more practical: Leaky relu. In *2020 IEEE Symposium on Computers and communications (ISCC)*, pages 1–7. IEEE, 2020.

[14] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[15] Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. Large-margin softmax loss for convolutional neural networks. *arXiv preprint arXiv:1612.02295*, 2016.

[16] Jun Han and Claudio Moraga. The influence of the sigmoid function parameters on the speed of backpropagation learning. In *International workshop on artificial neural networks*, pages 195–201. Springer, 1995.

[17] Babak Zamanlooy and Mitra Mirhassani. Efficient vlsi implementation of neural networks with hyperbolic tangent activation function. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 22(1):39–48, 2013.

[18] Zijun Zhang. Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pages 1–2. Ieee, 2018.

[19] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers*, pages 177–186. Springer, 2010.

[20] Long Wen, Liang Gao, Xinyu Li, and Bing Zeng. Convolutional neural network with automatic learning rate scheduler for fault classification. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12, 2021.

[21] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[22] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[23] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*, volume 2. Cambridge university press Cambridge, 2001.

[24] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019.

[25] Hai-Long Shi, Si-Yuan Liu, Xiao-Hui Wang, Wen-Li Yang, Zhan-Ying Yang, and Heng Fan. Coherence depletion in the grover quantum search algorithm. *Physical Review A*, 95(3):032307, 2017.

[26] Thomas E O'Brien, Brian Tarasinski, and Barbara M Terhal. Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments. *New Journal of Physics*, 21(2):023022, 2019.

[27] Yaakov S Weinstein, MA Pravia, EM Fortunato, Seth Lloyd, and David G Cory. Implementation of the quantum fourier transform. *Physical review letters*, 86(9):1889, 2001.

[28] Stephan Gulde, Mark Riebe, Gavin PT Lancaster, Christoph Becher, Jürgen Eschner, Hartmut Häffner, Ferdinand Schmidt-Kaler, Isaac L Chuang, and Rainer Blatt. Implementation of the deutsch–jozsa algorithm on an ion-trap quantum computer. *Nature*, 421(6918):48–50, 2003.

[29] Daniel Berrar et al. Cross-validation., 2019.

[30] Petro Liashchynskyi and Pavlo Liashchynskyi. Grid search, random search, genetic algorithm: a big comparison for nas. *arXiv preprint arXiv:1912.06059*, 2019.

[31] Mohamad Zaim Awang Pon and Krishna Prakash KK. Hyperparameter tuning of deep learning models in keras. *Sparklinglight Transactions on Artificial Intelligence and Quantum Computing (STAIQC)*, 1(1):36–40, 2021.

[32] Joseph O Ogutu, Torben Schulz-Streeck, and Hans-Peter Piepho. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In *BMC proceedings*, volume 6, pages 1–6. Springer, 2012.

[33] Jose M Alvarez and Mathieu Salzmann. Learning the number of neurons in deep networks, 2018.

[34] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3:1–40, 2016.

[35] Stuart Hadfield, Zhihui Wang, Bryan O'gorman, Eleanor G Rieffel, Davide Venturelli, and Rupak Biswas. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*, 12(2):34, 2019.

[36] Yaofeng Desmond Zhong, Tongtao Zhang, Amit Chakraborty, and Biswadip Dey. A neural ode interpretation of transformer layers. *arXiv preprint arXiv:2212.06011*, 2022.

[37] Geoffrey E Hinton, Alex Krizhevsky, and Sida D Wang. Transforming auto-encoders. In *Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21*, pages 44–51. Springer, 2011.

[38] Nikolas P Breuckmann and Xiaotong Ni. Scalable neural network decoders for higher dimensional quantum codes. *Quantum*, 2:68, 2018.

[39] Bei Li, Quan Du, Tao Zhou, Yi Jing, Shuhan Zhou, Xin Zeng, Tong Xiao, Jingbo Zhu, Xuebo Liu, and Min Zhang. Ode transformer: An ordinary differential equation-inspired model for sequence generation. *arXiv preprint arXiv:2203.09176*, 2022.

[40] John Charles Butcher. A history of runge-kutta methods. *Applied numerical mathematics*, 20(3):247–260, 1996.

[41] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient gan training. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

[42] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[43] Yifan Jiang, Shiyu Chang, and Zhangyang Wang. Transgan: Two pure transformers can make one strong gan, and that can scale up. *Advances in Neural Information Processing Systems*, 34:14745–14758, 2021.

[44] Alan C. Santos. O computador quântico da ibm e o ibm quantum experience. *Revista Brasileira de Ensino de Física*, 39(1), September 2016.

[45] Gil Kalai, Yosef Rinott, and Tomer Shoham. Google's quantum supremacy claim: Data, documentation, and discussion, 2023.

[46] Mariia Mykhailova. Teaching quantum computing using microsoft quantum development kit and azure quantum. In *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, September 2023.

[47] Justin A. Reyes, Dan C. Marinescu, and Eduardo R. Mucciolo. Simulation of quantum many-body systems on amazon cloud. *Computer Physics Communications*, 261:107750, April 2021.

[48] Vedran Dunjko and Hans J Briegel. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Reports on Progress in Physics*, 81(7):074001, 2018.

[49] Yagnik Chatterjee, Eric Bourreau, and Marko J Rančić. Solving various np-hard problems using exponentially fewer qubits on a quantum computer. *Physical Review A*, 109(5):052441, 2024.

[50] Martin Fürer. Solving np-complete problems with quantum search. In *Latin American Symposium on Theoretical Informatics*, pages 784–792. Springer, 2008.

[51] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[52] Jean-Baptiste Cordonnier, Andreas Loukas, and Martin Jaggi. On the relationship between self-attention and convolutional layers, 2020.

[53] Juyong Kim, Yookoon Park, Gunhee Kim, and Sung Ju Hwang. SplitNet: Learning to semantically split deep networks for parameter reduction and model parallelization. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1866–1874. PMLR, 06–11 Aug 2017.

[54] Basheer Qolomany, Majdi Maabreh, Ala Al-Fuqaha, Ajay Gupta, and Driss Benhaddou. Parameters optimization of deep learning models using particle swarm optimization. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 1285–1290, 2017.

[55] Chai Wah Wu. Prodsumnet: reducing model parameters in deep neural networks via product-of-sums matrix decompositions. *arXiv preprint arXiv:1809.02209*, 2018.

[56] Alireza Azadbakht, Saeed Reza Kheradpisheh, Ismail Khalfaoui-Hassani, and Timothée Masquelier. Drastically reducing the number of trainable parameters in deep cnns by inter-layer kernel-sharing. *arXiv preprint arXiv:2210.14151*, 2022.

[57] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021.