Continuous Variable Single Mode Quantum Decoder for Image Reconstruction and Denoising

Jasvith Raj Basani and Aranya Bhuti Bhattacherjee

Department of Physics, Birla Institute of Technology and Science, Pilani, Hyderabad Campus, Telangana 500078, India

(Dated: June 22, 2020)

Quantum computation using optical modes has been well-established in its ability to construct deep neural networks. We introduce a model that is the quantum analogue of the classical autoencoder - a neural network model that can reconstruct its input via dimensionality reduction and expansion through the phase-space formulation of quantum mechanics. The hallmark of the continuous-variable (CV) model is its ability to forge non-linear functions using a set of gates that allows it to remain completely unitary. We leverage this property of the CV model to encode and decode - classical information and demonstrate denoising applications using parallel single mode photonic circuits. The proposed model exemplifies that the appropriate photonic hardware can be integrated with present day optical communication systems to meet our information processing requirements. Here, using the Strawberry Fields software library on the MNIST dataset of handwritten digits, we demonstrate the adaptability of the network to learn classical information to fidelities of greater than 99.98%.

I. INTRODUCTION

Quantum Machine Learning [1] is an emerging field that has garnered interest from physicists and computer scientists alike, that aims at enhancing the methods used in machine learning using the non-classical effects of quantum mechanics, such as superposition or entanglement. Quantum computers offer an excellent platform to leverage these effects to explore variants of neural networks and their applications in machine learning.

Neural networks enjoy widespread success both in academia and the industry, with a plethora of applications ranging from solving simple classification problems to information security and processing. The advent of quantum technologies has resulted in what is being referred to as quantum-enhanced machine learning where we expect a speed-up either by employing genuine quantum effects, or by classical machine learning to improve quantum processes. A hybrid classical-quantum system achieves this speed-up by outsourcing computationally difficult subroutines to the integrated quantum device - specifically the quadratures of light in a quantum optical system.

The widely accepted model of the qubit based quantum computer has proven to be ill-suited to tackle continuous-valued problems [2]. Fortunately, the continuous-variable (CV) model proposes an alternative to the qubit - the quakmode, wherein information is encoded in the quantum states of bosonic modes - the ubiquity and features of which will be further discussed.

Noise in information systems has plagued the quality of signals and images and has vexed researchers from the days of development of telecommunication technology. The following work presents a hybrid classical-quantum optical system, inspired from the CV model that demonstrates denoising by layering single mode optical circuits.

This paper is organized as follows - in section II, we review the key concepts of classical autoencoders to arrive at the classical analogue of the loss function that will be used in the numerical experiments. We then introduce the key features of the continuous-variable model and its application in quantum computing in section III. We then move forward into the theoretical exposition of the quantum neural network model in section IV, followed by validating the theory by modelling numerical experiments in sections V and VI. Finally, after a discussion of the merits of the model, we conclude with the applications and future scope in this field.

II. THEORY OF CLASSICAL AUTOENCODERS

Autoencoders are a particular type of neural network that were first introduced in [3], trained in a manner consistent with the behaviour of identity function i.e. to replicate or reconstruct it’s input. Specifically, a section of the neural network compresses or encodes the input data into a more meaningful or useful representation, while the succeeding section decodes or reconstructs the output, keeping it as close to the input as possible by minimizing the loss due to dissimilarity. Hence, formally defined, the function to be learnt by the encoder $\phi_{in} : \mathbb{R}^{N_1} \rightarrow \mathbb{R}^{N_2}$ and the decoder $\phi_{out} : \mathbb{R}^{N_2} \rightarrow \mathbb{R}^{N_1}$, where $N_1 > N_2$ is:

$$\phi_{eq} = \arg\min_{\phi_{in} \phi_{out}} E\{L(x, \phi_{out} \circ \phi_{in}(x))\}$$  \hspace{1cm} (1)

where $E$ is the expectation value over $x$ and $L$ is the reconstruction loss. To understand and visualize the outputs of the trained encoder section, each hidden unit computes the function

$$z_i^{(n+1)} = \sigma(\sum_{j=1}^{m} W_{ij}^n x_j + b_i^n)$$
where the terms $\sigma, W$ and $b$ are the activation function, the weight matrix and the bias vectors respectively. Similarly, the output of the trained decoder section can be expressed as:

$$x'_{k}^{n+1} = \sigma'(\sum_{j=1}^{m} W_{kj} x_{j}^{n} + b_{k}^{n})$$

Hence, the reconstruction loss introduced in equation (1) can be solved as the norm or squared error between the input and output vector $x$.

$$L = ||x - x'||^2 = ||x - \sigma'(W'(\sigma(Wx + b)) + b'))||^2$$

The denoising autoencoder is a stochastic extension of the classical autoencoder - wherein the network tries to reconstruct the original image from a corrupted input by capturing statistical dependencies between the input and output. Further details on perspectives and methods of operation can be found in Ref. [16].

![Scaled architecture of the classical encoder network used in the following discussion - 784 units in the input layer, 512 and 64 units in the hidden layers, 2 units in the output layer, returning the $\hat{x}$ and $\hat{p}$ expectation values](image)

**FIG. 1.** Scaled architecture of the classical encoder network

III. QUANTUM COMPUTING WITH CONTINUOUS VARIABLES

As opposed to the conventional discrete set of coefficients which are used in the qubit expansions, the continuous variable (CV) model [4] uses a continuum of states as seen by the expansion of a qumode below:

$$|\psi\rangle = \int \psi(x)|x\rangle dx$$

where $|x\rangle$ are the eigenstates of the $\hat{x}$ quadrature, which will further be elaborated upon below. In the following sections, we reference the phase space formalism of the continuous variable quantum mechanics where we treat the conjugate variables $x$ and $p$ on equal footing which allows us to find symmetries with classical Hamiltonian dynamics.

The universality of the CV model has been well established and is based upon its ability to approximate a broad set of Hamiltonians that are polynomials, fixed in degree, which are functions of the $\hat{x}$ and $\hat{p}$ quadratures. In the discrete qubit system, we are enabled by a set of gates that allow any normalized system to transform through unitary operations, while in the CV model, we apply Gaussian and Non-Gaussian transformations to evaluate the the evolution of a state that takes the form $|\psi\rangle = e^{iHt}|0\rangle$ where $H$ is the generator or the Hamiltonian of the bosonic system and $|0\rangle$ is the vacuum state which we initialize to assume time dependant evolution. These properties will further be studied in the development of quantum neural networks. The $\hat{x}$ and $\hat{p}$ operators work over the domain of the entire real line as we see from the equations below:

$$\hat{x} = \int_{-\infty}^{+\infty} x|x\rangle\langle x|dx$$

$$\hat{p} = \int_{-\infty}^{+\infty} p|p\rangle\langle p|dp$$

The vectors $|x\rangle$ and $|p\rangle$ are orthogonal and form a continuous spectra, and are not normalizable since the trace is dependant on the basis chosen, with each satisfying $\langle x|x'\rangle = \delta(x - x')$. The operators are non-commutative as $[\hat{x}, \hat{p}] = i\Omega_{jk}$ are related by the Fourier transform and can be solved for using the Rotation Gaussian Gate listed below.

Rotation :

$$R(\theta) = e^{i\theta a^\dagger a}$$

Similarly, the set of other commonly used gates that are included in the model above are listed with their respective functions:

Interferometer :

$$U(\phi_{in}, \phi_{ex}) = BS(\frac{\pi}{4}, \frac{\pi}{2})(R(\phi_{in}) \otimes I)BS(\frac{\pi}{4}, \frac{\pi}{2})(R(\phi_{ex}) \otimes I)$$

Beamsplitter :

$$B(\theta, \phi) = e^{i\theta(e^{i\phi}a_{1}a_{2} + e^{-i\phi}a_{1}^{\dagger}a_{2}^{\dagger})}$$

Squeeze gate :

$$S(r, \phi) = e^{\frac{i}{2}(r e^{-i\phi}a_{2} - r e^{i\phi}a_{1}^{\dagger})}$$

Displacement gate :

$$D(\alpha) = e^{\alpha a^\dagger - \alpha^* a}$$

The CV formalism has long been established as an experimentally realizable and flexible scheme to operate upon, with previous literature covering optical systems [9 10], ion traps [11 12] and microwave systems [13 14]. The phase space formulation conventionally operates on
FIG. 2. The circuit model for a single layer of the quantum neural network comprising of the interferometer (composed of rotation and beamsplitter gates), squeeze gates, another interferometer, displacement and non-gaussian activation (which performs the non-linear transformation) gates as proposed in [5].

2N different real valued continuous variables \((x, p) \in \mathbb{R}^{2N}\). However, with a slight modification in labelling the observables, we can encode higher dimensional information into a single qumode - the observables being a set of matrices with the \(N^{th}\) diagonal element set to unity - giving us linear, hermitian and real-eigenvalued operators. Hence, the eigenstates of the \(n^{th}\) quadrature operator satisfies the following relations:

\[
\hat{n}|n\rangle = n|n\rangle \\
\langle n'|n\rangle = \delta(n - n') \\
\int |n\rangle\langle n| \, dn = 1
\]

IV. QUANTUM NEURAL NETWORKS

In the following section, we discuss the general scheme of a quantum neural network as presented in [5], as an analogue of the classical fully connected neural network. The output of the network can be described as the following:

\[
|x\rangle \rightarrow |\sigma(Wx + b)\rangle
\]

From figure 2, we see the implementation of a layer \(L\) of the CV based quantum neural network as proposed in [5]. As apparently presented, the function of the layer can be broken down as follows:

\[
L = \Phi \cdot D \cdot U_2 \cdot S \cdot U_1
\]  

Here, \(U_i\) is the \(i^{th}\) \(N\) mode interferometer, that can further be decomposed rectangularly or triangularly into combinations of beamsplitters and rotation gates, hence is a function of the polarization angle as well as the azimuthal rotation angle. Furthermore, \(S\) is the single mode squeeze gate, \(D\) is the single mode displacement gate and \(\Phi\) is the Non-Gaussian gate acting on a single mode, acting as non-linear activation function.

A powerful tool in linear algebra that can be leveraged is the singular value decomposition [24] or factorization of a matrix \(W\) into orthogonal matrices and a diagonal matrix, which is provided to us as an in-built structure in the combination of interferometers and squeeze gates. Hence, we can rewrite \(W\) as \(O_2 M O_1\) where \(O_i\) is the orthogonal matrix and \(M\) is the diagonal matrix. The following can be understood easily when described mathematically as:

\[
U_1(\theta_1, 0)|x\rangle = |O_1x\rangle \\
\otimes_{i=1}^N \, S(r_i)|x\rangle = |M O_1x\rangle \\
U_2(\theta_2, 0)|MO_1x\rangle = |O_2MO_1x\rangle = |Wx\rangle
\]

The bias and the non-linear activation function can be added to the transformed input via the displacement gate and a Non-Gaussian gate respectively, both shown such that the transformation in equation (2) is achieved.

\[
\otimes_{i=1}^N D(\alpha_i)|x\rangle = |x + b\rangle \\
\otimes_{i=1}^N \Phi(\lambda_i)|x\rangle = |\sigma(x)\rangle
\]

Deep learning models [25] are made by stacking layers of the aforementioned circuit end-to-end while varying the number of qumodes in each layer for the based on the amount of information encoded into each qumode.

V. QUANTUM DECODER

In the following section, we design a single mode quantum decoder, working in an orthogonal 28-dimensional
basis to demonstrate the reconstruction of an image from the MNIST dataset of handwritten digits [23]. We use a classical encoder scheme with the architecture as shown in figure [1]. The encoder compresses the 784 x 1 vector (reshaped from the 28 x 28 image) into 2 values - the expectation values of the single photon to be passed into the decoding network as shown in the figure below and employ 15 such layers with controllable parameters.

We define the normalized projection as follows, with the projection performed onto the subspace of the first 28 Fock states, corresponding to the size of the output image required.

$$|\psi_i\rangle = \hat{\Pi}_{28} |\Psi_i\rangle$$

This projected state is then modified as the network learns from the cost function given below, with the closeness between the target state and the learnt state measured by the fidelity defined by

$$C = \sum_{i=0}^{27} (|\langle i | \psi_i \rangle|^2 - 1)^2 + \gamma P(|\psi_i\rangle)$$

The plot in figure [5] shows us the characteristics of the learning curves for the network when the Adam [26] and RMSProp [27] Optimizers were used with the learning rate $\alpha$ set to 0.001.

FIG. 3. Single Mode Quantum Decoder Layer

FIG. 4. Inputs (left) and outputs (right) of the hybrid network - the images of the numbers '0' and '8' from the MNIST dataset of handwritten digits

FIG. 5. Plot of the cost function against iteration number for the Adam and RMSProp Optimizers used in training the network

VI. DENOISING QUANTUM DECODER

In the following section, we explore the aspect of denoising the input image with Additive White Gaussian
FIG. 6. Two parallel single mode quantum decoder layers, fed into the Oracle for storage and measurement - each learning the Fourier transforms of the noisy image and the noise.

Noise (AWGN). The Central Limit Theorem [15] establishes that the sum of independent random variables - in this case kinds of noise - tends towards a normal distribution, even if the original variables themselves are not normally distributed. AWGN is a model of noise, commonly used in information theory and processing so as to model the random effects of interactions that occur in nature. We leverage the property that this noise is white - implying that the power spectrum of the noise is a constant and that it is Gaussian - implying that the histogram has a normal distribution as shown in the figure below.

![Noise Histogram](image)

FIG. 7. Noise Histogram for Additive White Gaussian Noise with mean = 0.5 and standard deviation = 0.1

The proposed method for denoising uses a method similar to that described above, the difference being, in this case, as opposed to the previous case, where the images were learnt directly layer by layer, the states learnt are that of the 2-Dimensional Fourier Transform (2DFT) of the noisy images and a constant matrix, with all terms set to the mean of the noise. To execute this, we use 2 parallel single mode decoder layers, with the final learnt states being fed into the oracle that we introduce as \( \hat{J} \) followed by a 2-Dimensional Inverse Fourier Transform (2DIFT). The oracle \( \hat{J} \) essentially applies the displacement (difference) operator between the learnt states and acts as a quantum memory buffer [21] to store all the learnt states until they can be measured and the 2DIFT can be applied. Let \( C_i \) represent the cost function for the noisy image and \( C_n \) represent the cost function for the noise, with \( \phi^i(k) \) and \( \phi^n(k) \) representing the 2DFT of the noisy image and the noise respectively.

\[
C_i = \sum_{j=0}^{2^7} (|\langle j | \phi^i_j (k) \rangle|^2 - 1)^2
\]

\[
C_n = \sum_{j=0}^{2^7} (|\langle j | \phi^n_j (k) \rangle|^2 - 1)^2
\]

Each layer is fed an individual photon which has been encoded separately with with either the noisy image or the noise. Figure [8] illustrates the inputs and the outputs of the modified network.

![Input and Output](image)

FIG. 8. Input and Output of the Denoising Quantum Decoder - the first row illustrates denoising the number "3" from the MNIST dataset with AWGN noise (mean = 0.5, standard deviation = 0.1) filtered to give a mean squared error of approximately 1% while the second row illustrates the denoising of the number "6" from the MNIST dataset with AWGN noise (mean = 0.5, standard deviation = 0.2) filtered to give a mean squared error of approximately 4%

Colored images can be trifurcated into their respective
RGB values, each of which will be learnt separately via a 6-layer parallel single mode system. For the added purpose of demonstration, we have modified the architecture so as to present the denoised results of an image with the VIBGYOR colour spectrum.

The proposed architecture has demonstrated exemplar reconstruction and denoising capabilities using the quantum properties of light. Here, we see that a system as simple as a 6-layer single mode circuit could be used to denoise a coloured image with each layer learning its RGB components and noise respectively. Harnessing other properties of quantum mechanics - specifically, entanglement, would enable us to expand the basis of calculation thereby making this architecture flexible to encoding and decoding of information structured in any data-type. Optical quantum information processing presents with the added advantage that extremely large amounts of data can be handled at incredible speeds. Moreover, the no-cloning theorem ensures that states of any quantum mechanical system cannot be duplicated making it impossible to access information encoded into such a system.

To better understand the performance of our network on the denoising, we plot the average mean squared error against the standard deviation of the noise added to the original figure. We observe an increasing correlation between the noisy and denoised image due to the fact that the AWGN model is immune to frequency selectivity and is used to provide a simple tractable model to visualize the working of the network before increasing the complexity of the noise added.

The plot shows us that this architecture and network function well enough to remove noise approximately 30% in mean squared error units. For higher values of standard deviation of the AWGN, the clarity of the image drops rapidly, rendering the reconstructed image useless. Experimental realizations of such a network would expose the photon to a multitude of sources of noise - while the Central Limit Theorem ascertains that the sum of independent noise variables would result in a Normal distribution, there would be an abundance of frequency dependant noise as well. We envision that such filtering components can be incorporated without altering the simplicity of the cost function.

To our knowledge, this is the first time a single mode hybrid classical-quantum neural network has been presented in the literature. We have demonstrated a single mode hybrid classical-quantum neural network that leverages the quantum properties of electromagnetic fields for encoding and decoding classical information. The flexibility of such a network is a vastly enabling tool with applications ranging from quantum cryptography to scalable quantum communication networks. Photons play an essential role in
all quantum networking systems - either as information carriers or as mediators between quantum memories [21]. Integration with spin based repeater nodes [18, 19] for the prospects of a quantum internet [20] is a promising direction for future exploration. Another fruitful direction of research would be to amalgamate the other fundamental properties of quantum mechanics - specifically the uncertainty principle and entanglement to evaluate their role in determining the information capacity of a quantum neural network [32].

While single photon networks and gates are yet to be realized to implement quantum machine learning algorithms and long distance communication efficiently, we hope that with the advent of emerging photonic technologies such as lossless fiber optics and quantum technologies like semiconductor based spin interaction dependant nodal systems, such systems will be integrated into our lives seamlessly.

[1] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. Nature, 549(7671):195, 2017.
[2] Alejandro Perdomo-Ortiz, Marzio Bertini, John Realpe-Gonzalez, and Rupak Biswas. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. *arXiv:1708.09757*, 2017.
[3] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by error propagation. In Parallel Distributed Processing, Vol 1: Foundations. MIT Press, Cambridge, MA, 1986.
[4] Alessandro Ferraro, Stefano Oliva, and Matteo GA Paris. Gaussian states in continuous variable quantum information. *arXiv preprint quant-ph/0503237*, 2005.
[5] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd. Continuous-variable quantum neural networks, *arXiv:1806.06871* (2018).
[6] Nathan Killoran, Josh Izaac, Nicolas Quesada, Ville Bergholm, Matthew Amy, and Christian Weedbrook. Strawberry Fields: A Software Platform for Photonic Quantum Computing, Quantum, 3, 129 (2019).
[7] Thomas R. Bromley, Juan Miguel Arrazola, Soran Jahangiri, Josh Izaac, Nicolas Quesada, Alain Delgado Gran, Maria Schuld, Jeremy Swaninorton, Zeid Zabaneh, and Nathan Killoran. Applications of Near-Term Photonic Quantum Computers: Software and Algorithms, *arXiv:1912.07634* (2019).
[8] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafael Jozefowicz, Yangqing Jia, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernando Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
[9] Jun-ichi Yoshikawa, Shota Yokoyama, Toshiyuki Kaji, Chanond Sorphiphathphong, Yu Shiozawa, Kenzo Makino, and Akira Furusawa. Invited article: Generation of one-million-mode continuous-variable cluster state by unlimited time-domain multiplexing. *AIP Photonics*, 1(6):060801, 2016.
[10] Ulrik L Andersen, Jonas S Neergaard-Nielsen, Peter Van Loock, and Akira Furusawa. Hybrid discrete and continuous-variable quantum information. *Nature Physics*, 11(9):713, 2015.
[11] DM Meekhof, C Monroe, BE King, Wayne M Itano, and David J Wineland. Generation of nonclassical motional states of a trapped atom. *Physical Review Letters*, 76(11):1796, 1996.
[12] Chao Shen, Zhen Zhang, and L-M Duan. Scalable implementation of boson sampling with trapped ions. *Physical review letters*, 112(5):050504, 2014.
[13] K Moon and SM Girvin. Theory of microwave parametric down-conversion and squeezing using circuit QED. *Physical review letters*, 95(14):140504, 2005.
[14] Borja Peropadre, Gian Giacomo Guerricchi, Joonsuk Huh, and Alan Aspuru-Guzik. Proposal for microwave boson sampling. *Physical review letters*, 117(14):140505, 2016.
[15] Bray, Inre; Vu, Van (2007). "Central limit theorems for Gaussian polytopes". *Annals of Probability*. Institute of Mathematical Statistics. 35 (4): 15931621. *arXiv:math/0610192*. doi:10.1214/009117906000000791.
[16] P. Vincent, H. Larochelle Y. Bengio and P.A. Manzagol, Extracting and Composing Robust Features with Denoising Autoencoders, Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML08), pages 1096 - 1103, ACM, 2008.
[17] A. Mari, T. R. Bromley, J. Izaac, M. Schuld, and N. Killoran, Transfer learning in hybrid classical-quantum neural networks, *arXiv:1912.08278* [quant-ph, stat], Dec, 2019.
[18] Steinbrecher, G. R., J. P. Olson, D. Englund, and J. Carolan, Quantum optical neural networks, *arXiv:1808.10047*, 2018.
[19] Atatüre, M., Englund D., Vanvickas N., Lee S.Y. & Wrachtrup J, "Material platforms for spin-based photonic quantum technologies," Nat. Rev. Mater. 3, 3851 (2018).
[20] Kimble, H. J., The quantum internet, Nature (London) 453, 1023, 2008.
[21] A. I. Lvovsky, B. C. Sanders, and W. Tittel. Optical quantum memory. *Nature Phot.*, 3:713, 2015.
[22] S. Slussarenko and G.J. Pryde, Photonic quantum information processing: A concise review. *Applied Physics B: Lasers and Optics*, 123(9):1-23, 2017.
[23] M. Atre, et al. “Material platforms for spin-based photonic quantum technologies,” *Nat. Rev. Mater.* 3, 3851 (2018).
[24] S. Slussarenko and G.J. Pryde, Photonic quantum information processing: A concise review. *Applied Physics Reviews* 6, 041303 (2019).
[25] R. Annavarapu, Singular Value Decomposition and the Centrality of Lwvin Orthogonalizations. *American Journal of Computational and Applied Mathematics* 3, 33 (2013).
[26] Deep Learning (Ian J. Goodfellow, Yoshua Bengio and Aaron Courville), MIT Press, 2016.
[26] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In ICLR, 2015

[27] Hinton N. Srivastava and K. Swersky. Neural networks for machine learning-lecture 6a-overview of mini-batch gradient descent, 2012

[28] Zhou, Y., Jiao, J., Huang, H., Wang, Y., Wang, J., Shi, H., Huang, T.: When AWGN-based denoiser meets real noises. arXiv preprint arXiv:1904.03485 (2019)

[29] Knill, E., R. Laflamme, and G. J. Milburn, 2001, Nature 409, 46. Kok, P., and B. Lovett, 2010, Introduction to optical quantum information processing (Cambridge University Press, Cambridge).

[30] John Preskill, Quantum Computing in the NISQ era and beyond, (2018), arXiv:1801.00862

[31] W.K. Wootters and W.H. Zurek, ‘A Single Quantum Cannot be Cloned’, Nature 299, t02-803 (1982).

[32] L. G. Wright and P. L. McMahon, The capacity of quantum neural networks, (2019), arXiv:1908.01364