

Quantum Layers for Deep Neural Network

Tom Smit, Guy Puts, Vincenzo Lipardi, Niels Neumann, Kurt Driessens, and Domenica Dibeneditto

Department of Advanced Computing Sciences, Maastricht University, The Netherlands

Department of Applied Cryptography and Quantum Algorithms, TNO, The Netherlands

`domenica.dibeneditto@maastrichtuniversity.nl`

Integrating quantum computation into deep neural networks remains a central challenge in quantum machine learning, because quantum measurements are probabilistic and do not naturally fit the gradient-based optimization pipelines used in classical deep learning. This work investigates trainable quantum layers as modular representation-learning components that can be inserted into pre-trained classical neural networks without requiring end-to-end quantum backpropagation. We introduce two architectures. The first is a circuit-based quantum layer in which a parameterized quantum circuit acts as a latent-space transformer. Its architecture is optimized automatically using Progressive Widening Monte Carlo Tree Search, guided by task-specific cost functions. The second is a trainable Open Quantum System (OQS) layer, where learning arises from a compact system–environment interaction rather than from a deep variational quantum circuit. In this model, fixed Pauli-basis dynamics are replaced by trainable interaction operators to increase expressivity. Both layers are evaluated in AutoEncoder-based anomaly detection pipelines, where the quantum layer is inserted at or after the bottleneck and the transformed latent representations are assessed using an Isolation Forest or OQS-based classifier. Results on benchmark datasets show that quantum transformations can improve latent-space separability and anomaly detection performance compared with classical baselines. The circuit-based layer achieves the strongest improvement when the search is guided by a triplet-loss objective, while the OQS layer provides a compact alternative with lower architectural complexity and improved performance when additional trainable rotations are introduced. These findings suggest that carefully designed quantum transformations can enrich classical representations and provide a practical route toward scalable hybrid quantum–classical deep learning.

Keywords: Quantum Machine Learning, Quantum Network Layer, Deep Learning, Quantum Open Systems, Quantum Architecture Search.

1 Trainable Quantum Layers for AutoEncoder-Based Anomaly Detection

Motivation and contributions. Quantum Machine Learning explores how quantum computation can enhance machine learning in terms of learning, representation, or generalization [1, 2]. In the current noisy intermediate-scale quantum (NISQ) era [3], this typically motivates hybrid architectures that couple quantum components with classical neural networks. A key challenge is integrating quantum layers into deep learning pipelines. While classical networks rely on backpropagation, quantum measurements yield probabilistic classical outcomes, making direct gradient propagation through measured quantum components non-trivial. Parameter-shift rules can train variational quantum circuits, but they introduce additional computational costs and may become difficult to scale with increasing circuit depth. Hybrid classical–quantum neural networks have been studied by replacing neurons, layers, or convolutional components with variational quantum circuits [4, 5]. Although promising, these approaches often depend on manually designed circuits, additional trainable quantum gates, parameter-shift optimization, and deeper architectures.

This work studies two trainable quantum-layer architectures inserted into AutoEncoder-based anomaly detection (AD) pipelines. The first is a circuit-based layer whose parameterized quantum circuit is optimized automatically through Progressive Widening Monte Carlo Tree Search (PWMCTS) [8], using reconstruction loss, inter-class mean separation, and triplet loss as task-specific objectives. The second is an Open Quantum System (OQS)-based layer, where latent features are encoded into the amplitudes of an environment state [13] and mapped to a quantum system through a unitary interaction with trainable rotation parameters. The main contributions are: (i) two trainable quantum-layer architectures for deep neural networks; (ii) a gradient-free quantum architecture search pipeline for PQC layers tailored to latent-space structuring; (iii) an AutoEncoder-based anomaly-detection framework for evaluating quantum transformations of latent representations; and (iv) empirical evidence, under classical simulation, that quantum layers can improve latent-space separability and anomaly-detection performance.

Architecture-searched circuit layer. The first approach uses a parameterized quantum circuit (PQC) as a trainable layer acting on the latent representation of a classical AutoEncoder (AE). This idea was first explored by Sakhnenko et al. [7], who showed that the performance of a hybrid classical–quantum AutoEncoder depends strongly on the chosen PQC architecture. Their analysis related circuit descriptors, such as the number of entangling layers, the Meyer–Wallach measure, and Sim expressivity, to performance metrics such as precision, recall, and F1-score. Although this provides useful design intuition, it still requires substantial manual circuit engineering. To reduce this

manual design step, we use Progressive Widening Monte Carlo Tree Search (PWMCTS) [8] to search for suitable PQC architectures. PWMCTS is a gradient-free Quantum Architecture Search method in which each node of the search tree represents a candidate circuit ansatz, and each edge corresponds to a circuit-modification action. These actions include adding, removing, replacing, or modifying gates. The search is applied to the variational block of the circuit, while the data-encoding and measurement layers are kept fixed. Candidate circuits are evaluated through task-specific latent-space objectives: reconstruction loss, inter-class mean separation, and triplet loss. The complete pipeline is shown in Fig. 1a. The transformed latent vectors are used as input to the anomaly-detection stage. In this work, the downstream detector is an Isolation Forest, which evaluates whether the quantum-transformed latent space improves the separation between normal and anomalous samples. In this way, PWMCTS is not used only to search for expressive circuits, but to identify circuit architectures that improve the structure of the AutoEncoder latent space for the target anomaly-detection task. After convergence, or after the iteration budget is reached, the best ansatz found by PWMCTS is combined with the fixed encoding and measurement layers to form the final circuit-based quantum layer. The resulting PQC defines a learned transformation of the AutoEncoder bottleneck, and its usefulness is evaluated by the anomaly-detection performance obtained from the transformed latent representations.

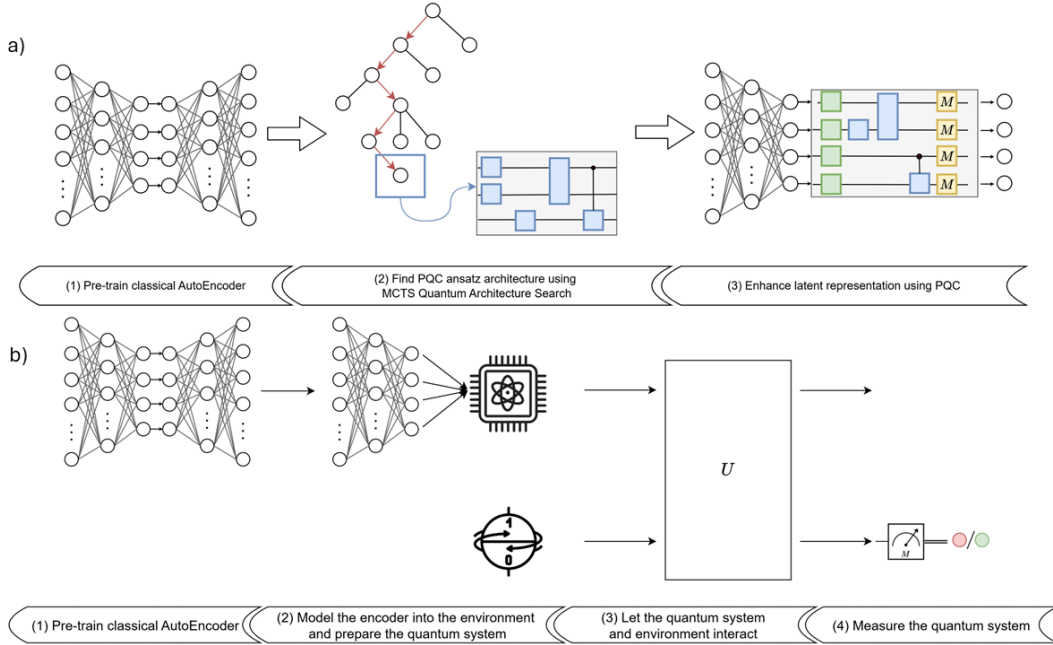


Figure 1: Pipelines: (a) PWMCTS-based circuit layer: a pre-trained AutoEncoder maps each sample to a latent vector, which is encoded into a quantum state and processed by a PQC whose ansatz is selected through PWMCTS. Optional data re-uploading alternates encoding and variational blocks, and final Z-basis measurements yield the quantum-transformed latent representation. (b) OQS-based layer: the latent vector is encoded as an environment state and processed through a trainable system-environment interaction, producing an anomaly score from the reduced system state.

Open Quantum System-based layer. The second approach uses an Open Quantum System (OQS)-inspired quantum layer. In contrast to closed quantum systems, open systems evolve through interaction with an environment. Such interactions are usually associated with decoherence and dissipation, but they can also be interpreted as structured information exchange between a system and its surroundings. In this work, we use an Open Quantum System (OQS)-inspired layer in which the AutoEncoder latent vector is encoded as an environment state and processed through a trainable system-environment interaction. After pre-training the AutoEncoder, each input x_i is mapped to a latent vector $z_i = E(x_i)$. The normalized latent vector is then encoded into an environment state $\rho_E(z_i)$. A single system qubit, initialized in a fixed superposition state, interacts with this environment through

$$U(\Theta, W) = \exp[i\sigma_S(\Theta) \otimes \sigma_E(W)],$$

where W contains feature-dependent interaction weights and $\Theta = (\theta, \phi, \lambda)$ parameterizes a trainable rotated system operator. The reduced system state is then obtained as

$$\rho_{\text{out}}(z_i) = \text{Tr}_E[U(\Theta, W)(\rho_S \otimes \rho_E(z_i))U^\dagger(\Theta, W)],$$

and the diagonal element $p_1(z_i) = \langle 1 | \rho_{\text{out}}(z_i) | 1 \rangle$ is used as the anomaly probability. The complete pipeline is shown in Fig. 1b. This formulation concentrates learning in a compact interaction operator rather than in a deep variational circuit, providing a parameter-efficient alternative to the PWMCTS-based layer.

2 Experimental protocol and results

The layers are evaluated on AD. After autoencoder pre-training, the encoder is frozen and the bottleneck features are transformed by either the PWMCTS circuit layer or the OQS interaction layer. The circuit-transformed features are passed to an Isolation Forest detector [14], while the OQS layer produces a binary decision from the measured system state. Performance is measured using precision, recall, and F1-score against the classical autoencoder baseline. The circuit experiments use the Statlog satellite dataset converted into a binary anomaly-detection task by grouping minority classes as anomalies [15, 16]. The OQS experiments use both Statlog and Arrhythmia [17].

Table 1: AD performance after inserting a quantum layer at the autoencoder bottleneck. HAE denotes the hybrid autoencoder explained in Section 1 with a PWMCTS-designed circuit layer; OQS denotes the IQC-AIL layer with column-wise normalization.

Dataset	Model / reward	Precision	Recall	F1-score
Statlog	Classical AE baseline	0.54	0.58	0.56
Statlog	HAE, reconstruction reward	0.56	0.61	0.59
Statlog	HAE, mean-separation reward	0.58	0.65	0.61
Statlog	HAE, triplet-loss reward	0.63	0.69	0.66
Statlog	OQS / IQC-AIL	0.601	0.668	0.628
Arrhythmia	Classical AE baseline	0.63	0.72	0.68
Arrhythmia	OQS / IQC-AIL	0.603	0.468	0.521

Table 1 shows that all PWMCTS objectives improve over the classical autoencoder baseline on Statlog. Reconstruction-based search gives the smallest gain, whereas mean separation and triplet loss produce stronger improvements. The triplet-loss reward achieves the best result, increasing the F1-score from 0.56 to 0.66. This indicates that the search objective should be aligned with the geometry required by the downstream anomaly detector rather than with reconstruction fidelity alone. The OQS layer achieves competitive performance on Statlog, with an F1-score of 0.628, but lower performance on Arrhythmia, suggesting that the current interaction model remains limited on more diffuse or higher-dimensional latent distributions. Taken together, the results identify two routes toward trainable quantum layers: PWMCTS addresses task-adapted ansatz design, while the OQS formulation addresses trainability and parameter compactness. It should be said however that the experiments are of such a small scale, that a quantum computer will make no positive difference. Scaling the circuit size up will make the circuits untrainable very fast due to known issues for PQC’s, given the current state of quantum hardware. This introduces a fundamental mismatch between the compact representation learning of autoencoders, and the need for large problems when searching for a possible quantum speedup. Additional challenges might be discovered when deploying quantum layers at larger scale.

3 Conclusion

The results show that trainable quantum layers can reshape autoencoder latent spaces to improve AD. For circuit-based layers, performance depends strongly on the architecture-search objective: latent-geometry rewards, especially triplet loss, outperform reconstruction-based rewards. This suggests that quantum-layer design should be driven by the downstream task rather than reconstruction fidelity alone. The OQS layer offers a complementary perspective. Although more dataset-dependent, it shows that useful trainable quantum transformations need not rely on deep variational circuits. By concentrating learning in compact system–environment interactions, OQS layers offer a route toward lower optimization complexity and parameter-efficient quantum representations. Overall, we presented a controlled framework for studying quantum transformations as modular representation-learning components. While circuit-based layers can increase expressive power in some regimes, their integration into neural networks quickly faces scalability limits. Future work should therefore focus on hybrid solvers for scalability, and on richer encodings, more expressive interaction operators, larger benchmarks, and hardware-aware implementations for fundamental QML research.

References

- [1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, pp. 195–202, 2017.
- [2] M. Schuld, I. Sinayskiy, and F. Petruccione, “An introduction to quantum machine learning,” *Contemporary Physics*, vol. 56, no. 2, pp. 172–185, 2015.
- [3] J. Preskill, “Quantum computing in the NISQ era and beyond,” *Quantum*, vol. 2, p. 79, 2018.
- [4] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, “Parameterized quantum circuits as machine learning models,” *Quantum Science and Technology*, vol. 4, no. 4, p. 043001, 2019.
- [5] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, “Variational quantum algorithms,” *Nature Reviews Physics*, vol. 3, pp. 625–644, 2021.
- [6] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Communications*, vol. 9, p. 4812, 2018.
- [7] A. Sakhnenko, C. O’Meara, K. J. B. Ghosh, C. B. Mendl, G. Cortiana, and J. Bernabe-Moreno, “Hybrid classical-quantum autoencoder for anomaly detection,” *Quantum Machine Intelligence*, vol. 4, p. 27, 2022.
- [8] V. Lipardi, D. Dibenedetto, G. Stamoulis, and M. H. M. Winands, “Quantum circuit design using a Progressive Widening enhanced Monte Carlo Tree Search,” *Advanced Quantum Technologies*, vol. 8, no. 10, 2025.
- [9] P.-Y. Wang, M. Usman, U. Parampalli, L. C. L. Hollenberg, and C. R. Myers, “Automated quantum circuit design with nested Monte Carlo Tree Search,” arXiv:2207.00132, 2022.
- [10] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015.
- [11] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*. Oxford, U.K.: Oxford University Press, 2002.
- [12] J. Zhang, Z. Li, R. He, J. Zhang, B. Wang, Z. Li, and T. Niu, “Interactive quantum classifier inspired by quantum open system theory,” in *Proceedings of the 2021 International Joint Conference on Neural Networks*, pp. 1–7, 2021.
- [13] E. B. Brito, F. M. de Paula Neto, and N. K. Bernardes, “Quantum classifier based on open quantum systems with amplitude information loading,” *Quantum Information Processing*, vol. 23, p. 338, 2024.
- [14] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *Proceedings of the 2008 IEEE International Conference on Data Mining*, pp. 413–422, 2008.
- [15] M. Kelly, R. Longjohn, and K. Nottingham, “The UCI Machine Learning Repository,” University of California, Irvine, 2024. Available: <https://archive.ics.uci.edu>
- [16] A. Srinivasan, “Statlog (Landsat Satellite),” UCI Machine Learning Repository, 1993. Available: <https://archive.ics.uci.edu/dataset/146/statlog+landsat+satellite>
- [17] H. Guvenir, B. Acar, G. Demiroz, A. Cekin, and R. Quinlan, “Arrhythmia,” UCI Machine Learning Repository, 1998. Available: <https://archive.ics.uci.edu/dataset/5/arrhythmia>