Enhancing Quantum Diffusion Models for Complex Image Generation

Jeongbin Jo[1], Santanam Wishal[2], Shah Md Khalil Ullah[3], Shan Kowalski[4], and Dikshant Dulai[5]

[1]Department of Physics , Yonsei University , *jeongbin033@yonsei.ac.kr*
[2]Asia Cyber University , *santawishal17@gmail.com*
[3]Khulna University of Engineering and Technology , *hamimkhandakar222@gmail.com*
[4]Eleven Dimension LLC , *shanjdk2012@gmail.com*

January 30, 2026

Abstract

Quantum generative models offer a novel approach to exploring high-dimensional Hilbert spaces but face significant challenges in scalability and expressibility when applied to multi-modal distributions. In this study, we explore a **Hybrid Quantum-Classical U-Net** architecture integrated with **Adaptive Non-Local Observables (ANO)** as a potential solution to these hurdles. By compressing classical data into a dense quantum latent space and utilizing trainable observables, our model aims to extract non-local features that complement classical processing. We also investigate the role of **Skip Connections** in preserving semantic information during the reverse diffusion process. Experimental results on the full MNIST dataset (digits 0-9) demonstrate that the proposed architecture is capable of generating **structurally coherent and recognizable images** for all digit classes. While hardware constraints still impose limitations on resolution, our findings suggest that hybrid architectures with adaptive measurements provide a **feasible pathway** for mitigating mode collapse and enhancing generative capabilities in the NISQ era.

Contents

1 Intrduction

Quantum Machine Learning (QML) is an emerging interdisciplinary research field that integrates principles of quantum computing with classical machine learning techniques to enhance data processing, pattern recognition, and optimization tasks. By exploiting quantum mechanical phenomena such as superposition, entanglement, and quantum interference, QML aims to provide computational advantages over classical learning algorithms, particularly for high-dimensional and complex datasets.

In classical machine learning, the performance of algorithms is often constrained by computational complexity and memory requirements as data size and model depth increase. Quantum computing, on the other hand, offers a fundamentally different computation paradigm, where information is represented using quantum bits (qubits) that can exist in multiple states simultaneously. This property allows quantum algorithms to explore large solution spaces more efficiently than their classical counterparts in certain scenarios.

One of the key motivations behind QML is the potential for speedup in tasks such as linear algebra operations, optimization, and sampling, which form the backbone of many machine learning models. Quantum-enhanced algorithms, including quantum support vector machines, variational quantum circuits, and quantum neural networks, have been proposed to accelerate learning processes and improve model expressiveness.

Despite its promising advantages, QML also faces several challenges. Current quantum hardware, commonly referred to as Noisy Intermediate-Scale Quantum (NISQ) devices, suffers from limited qubit counts, short coherence times, and gate errors. These hardware constraints restrict the depth and scalability of quantum models, often requiring hybrid quantum-classical approaches for practical implementations. Additionally, the theoretical quantum advantage of many QML algorithms remains an open research question, as classical algorithms continue to improve rapidly.

In summary, Quantum Machine Learning represents a promising yet evolving research direction. While it offers the potential for enhanced computational power and novel learning capabilities, significant challenges related to hardware limitations, noise, and algorithmic scalability must be addressed before widespread real-world adoption becomes feasible.

1.1 Quantum Convolutional Neural Network

2 Diffusion Model

2.1 Classical Diffusion Model

Diffusion models are a class of generative models in machine learning inspired by non-equilibrium thermodynamics. The goal is to learn a diffusion process that destroys structure in data, and then learn a reverse process that restores structure to generate new data samples from noise.[3]

2.1.1 Forward Process: Diffusion Process

The forward process is defined by a Markov chain that gradually adds Gaussian noise to the data according to a variance schedule $\beta_1, \ldots, \beta_T$.

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}\right) \tag{1}$$

Using the notation $\alpha_t = 1-\beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, we can sample x_t at any arbitrary time step t directly from x_0 in closed form:

$$q(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}\right) \tag{2}$$

This property allows for efficient training by eliminating the need to iterate through all previous timesteps.

2.1.2 Reverse Process: Generative Process

The generative process[5] is defined as the reverse Markov chain. Since the exact reverse posterior $q(x_{t-1}|x_t)$ is intractable, we approximate it using a neural network with parameters θ:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\right) \tag{3}$$

Here, $\mu_\theta(x_t, t)$ is the predicted mean, and $\Sigma_\theta(x_t, t)$ is the covariance (often fixed to $\beta_t\mathbf{I}$ or learned).

2.1.3 Loss Function

Training is performed by optimizing the variational lower bound (VLB) on the negative log-likelihood.

$$\mathcal{L} = \mathbb{E}\left[-\log p_\theta(\mathbf{x}_0)\right] \leq \mathbb{E}_q\left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right] \tag{4}$$

This can be rewritten as a sum of KL divergence[8] terms:

$$\mathcal{L} = \mathbb{E}_q\left[\underbrace{D_{KL}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))}_{L_T} + \sum_{t>1}\underbrace{D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}}\underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}\right] \tag{5}$$

In practice, Ho et al. found that a simplified objective function yields better sample quality. The objective is to predict the noise ϵ added to x_0:

$$\mathcal{L}_{\text{simple}}(\theta) = \mathbb{E}_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\|^2\right] \tag{6}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ and ϵ_θ is a function approximator (e.g., U-Net).

2.1.4 Continuous-Time Formulation: SDE

The discrete diffusion process can be generalized to continuous time using Stochastic Differential Equations (SDEs). The forward process is described by the following Itô SDE[7]:

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w} \tag{7}$$

where $\mathbf{f}(\cdot, t)$ is the drift coefficient, $g(t)$ is the diffusion coefficient, and $\mathbf{w}$ is a standard Wiener process.

Remarkably, the reverse process—generating data from noise—is also a diffusion process governed by the *reverse-time SDE*[1]:

$$d\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] dt + g(t)d\bar{\mathbf{w}} \tag{8}$$

Here, dt represents a negative time step, and $\bar{\mathbf{w}}$ is the Brownian motion in reverse time. The term $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$, known as the *score function*, points towards high-density regions of the data.

Therefore, the core task of the diffusion model is to learn a score-based model $s_\theta(\mathbf{x}, t) \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ using a neural network (or PQC in our case), allowing us to numerically solve the reverse SDE to generate samples from noise.

2.2 Quantum Diffusion Model

While classical diffusion models operate on probability distributions over classical data, Quantum Diffusion Models (QDMs) extend this concept to the Hilbert space of quantum states. The goal is to generate quantum states (density matrices) from a maximally mixed state.

2.2.1 Forward Process: Depolarizing Channel

Instead of adding Gaussian noise, the forward process in QDM is typically modeled as a standard depolarizing channel acting on a density matrix ρ. For a system of n qubits with dimension $d = 2^n$, the state at step t is given by:

$$\rho_t = \mathcal{E}_t(\rho_{t-1}) = (1 - p_t)\rho_{t-1} + p_t \frac{I}{d} \tag{9}$$

where $p_t \in [0, 1]$ is the depolarization probability (noise schedule). Similar to the classical case, we can express ρ_t directly from the initial state ρ_0. Let $\alpha_t = \prod_{s=1}^{t}(1 - p_s)$, then:

$$\rho_t = \alpha_t \rho_0 + (1 - \alpha_t)\frac{I}{d} \tag{10}$$

As $t \to T$, $\alpha_T \to 0$, and the state converges to the maximally mixed state $\rho_T \approx \frac{I}{d}$, which contains no information about ρ_0.

2.2.2 Reverse Process: Quantum Denoising

The reverse process aims to restore the quantum state from the noise. This is modeled by a parameterized quantum circuit (PQC), denoted as a unitary operator $U(\theta)$. The discrete reverse step can be approximated as:

$$\rho_{t-1} \approx \mathcal{D}_\theta(\rho_t, t) = U(\theta_t)\rho_t U^\dagger(\theta_t) \tag{11}$$

For more complex generative tasks, the reverse process may involve ancillary qubits and measurements to simulate non-unitary maps.

2.2.3 Loss Function: Fidelity or Trace Distance

Unlike the KL divergence used in classical models, quantum models often use Fidelity or Hilbert-Schmidt distance to measure the closeness between the generated state and the target state. A common objective is to maximize the overlap with the target pure state $|\psi\rangle$:

$$\mathcal{L}(\theta) = 1 - \mathbb{E}\left[\langle\psi|\,\rho_{\text{gen}}(\theta)\,|\psi\rangle\right] \tag{12}$$

Alternatively, for density matrices, we minimize the trace distance or maximize the quantum fidelity $F(\rho, \sigma) = (\text{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2$.

3 Quantum Data Encoding

We referred to IBM Quantum Platform, especially Quantum Machine Learning.[6]

3.1 Basis Encoding

Basis encoding maps a classical P-bit string directly to a computational basis state of a P-qubit system. For a single feature represented as binary bits $(b_1, b_2, \ldots, b_P)$, the quantum state is:

$$|x\rangle = |b_1, b_2, \ldots, b_P\rangle\,, \quad b_i \in \{0, 1\} \tag{13}$$

3.2 Amplitude Encoding

Amplitude encoding maps a normalized classical N-dimensional data vector $\vec{x}$ to the amplitudes of an n-qubit quantum state, where $n = \lceil \log_2 N \rceil$.

$$|\psi_x\rangle = \frac{1}{\alpha}\sum_{i=1}^{N} x_i\,|i\rangle \tag{14}$$

Here, $|i\rangle$ is the computational basis state and α is a normalization constant ensuring $\langle\psi_x|\psi_x\rangle = 1$. And α is a normalization constant to be determined from the data being encoded. We use amplitude encoding in the quantum diffusion model.

$$\alpha = \sqrt{\sum_{i=1}^{N} |x_i|^2} \tag{15}$$

3.3 Angle Encoding

Angle encoding maps each feature x_k to the rotation angle of a qubit using R_Y gates. For a data vector $\vec{x}$, the state is a product state:

$$|\vec{x}\rangle = \bigotimes_{k=1}^{N} R_Y(x_k) |0\rangle = \bigotimes_{k=1}^{N} \left(\cos\left(\frac{x_k}{2}\right) |0\rangle + \sin\left(\frac{x_k}{2}\right) |1\rangle \right) \tag{16}$$

3.4 Phase Encoding

Phase encoding maps data features to the phase of qubits using Phase gates $P(\phi)$, typically applied after Hadamard gates.

$$|\vec{x}\rangle = \bigotimes_{k=1}^{N} P(x_k) |+\rangle = \frac{1}{\sqrt{2^N}} \bigotimes_{k=1}^{N} \left(|0\rangle + e^{ix_k} |1\rangle \right) \tag{17}$$

3.5 Dense Angle Encoding

Dense angle encoding encodes two features, x_k and x_l, into a single qubit using both Y-axis and Z-axis rotations.

$$|x_k, x_l\rangle = R_Z(x_l) R_Y(x_k) |0\rangle = \cos\left(\frac{x_k}{2}\right) |0\rangle + e^{ix_l} \sin\left(\frac{x_k}{2}\right) |1\rangle \tag{18}$$

Extending this to more features, the data vector $\vec{x} = (x_1, ..., x_N)$ can be encoded as:

$$|\vec{x}\rangle = \bigotimes_{k=1}^{N/2} \left(\cos x_{2k-1} |0\rangle + e^{ix_{2k}} \sin x_{2k-1} |1\rangle \right) \tag{19}$$

4 Quantum Circuit

The choice of quantum circuit ansatz is critical in designing Quantum Machine Learning (QML) models, as it determines the expressivity and trainability of the network within the constraints of Noisy Intermediate-Scale Quantum (NISQ) devices. Based on recent surveys and specific applications in diffusion models, we categorize relevant ansatzes into general-purpose architectures and task-specific designs for image processing.

4.1 General Variational Ansatzes

- **Hardware-Efficient Ansatz (HEA):** Designed to minimize circuit depth and gate count on NISQ devices, HEA utilizes a layered structure of parameterized single-qubit rotations followed by entangling gates (e.g., CNOTs) tailored to the specific connectivity of the quantum hardware [4]. While it offers high implementability, it is known to suffer from barren plateaus if not carefully initialized.

- **Hamiltonian Variational Ansatz (HVA):** Inspired by the Quantum Approximate Optimization Algorithm (QAOA) and adiabatic quantum computation, HVA constructs the ansatz based on the problem Hamiltonian. It is particularly effective for physics-inspired problems but introduces a hierarchical structure to manage complexity [4].

4.2 Ansatzes for Quantum Machine Learning and Vision

For high-dimensional data such as images, specialized ansatzes inspired by tensor networks and convolutional neural networks are preferred.

- **Quantum Convolutional Neural Network (QCNN):** QCNN adapts the structure of classical CNNs to the quantum domain. It consists of alternating convolutional layers (quasilocal unitary evolutions) and pooling layers (measurements or partial traces) to hierarchically reduce dimensionality while extracting features [4]. This structure effectively captures spatial correlations, making it suitable for image generation tasks.

- **Multiscale Entanglement Renormalization Ansatz (MERA):** Originally designed for simulating quantum many-body systems, MERA represents quantum states using a hierarchical tensor network. It efficiently captures correlations at different length scales, which is analogous to the feature extraction process in deep learning [4].

4.3 Optimal General Two-Qubit Ansatz (Vatan-Williams Decomposition)

We adopt the optimal quantum circuit decomposition for general two-qubit interactions as proposed by Vatan and Williams [14]. Specifically, we utilize the explicit circuit construction for the non-local interaction block $N(\alpha, \beta, \gamma)$, which is the core component of the ansatz.

The circuit structure, depicted in Fig. 1, consists of three CNOT gates interleaved with parameterized single-qubit rotations. This configuration is capable of simulating the unitary operator $U = \exp(-i(\alpha\sigma_x \otimes \sigma_x + \beta\sigma_y \otimes \sigma_y + \gamma\sigma_z \otimes \sigma_z))$ up to local basis transformations.

The parameters α, β, γ serve as the learnable weights of the quantum diffusion model, allowing the network to capture complex correlations between features represented by qubits q_0 and q_1.

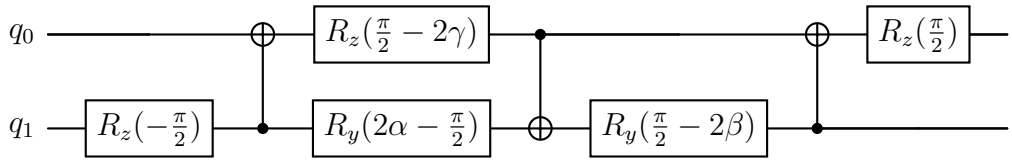

Figure 1: The parameterized quantum circuit used in our Quantum Bottleneck. This structure implements the optimal decomposition for arbitrary two-qubit entangling gates, characterized by the parameters α, β, γ. Note the alternating CNOT direction and the specific rotation angles derived from the canonical decomposition.

4.4 Global Information Propagation: 1D Cluster State Mixing

While the convolutional layer handles local correlations, a mechanism is required to propagate information globally across the qubit register. To achieve this efficiently, we employ a *Phase Mixing* layer inspired by 1D Cluster States (or Graph States) [12].

This layer, U_{Mix}, consists of three steps:

1. **Superposition Initialization:** A layer of Hadamard gates ($H^{\otimes N}$) transforms the basis states into a uniform superposition, maximizing the parallel processing capability.
2. **Linear Entanglement Chain:** We apply a sequence of Controlled-Phase (CZ) gates between adjacent qubits. This creates a 1D cluster state backbone, establishing a linear graph of entanglement that serves as a channel for information flow.
3. **Spatial Adaptivity:** Finally, a layer of parameterized rotation gates $R_X(\phi_i)$ is applied to each qubit. Unlike global mixing strategies (e.g., Grover Mixer) that apply a uniform operation, these individual rotations allow the model to learn spatially adaptive weights, emphasizing specific regions of the latent space (e.g., distinguishing object boundaries from background).

Mathematically, the mixing unitary is expressed as:

$$U_{\text{Mix}}(\phi) = \left(\bigotimes_{j=1}^{N} R_X(\phi_j) \right) \left(\prod_{j=1}^{N-1} CZ_{(j,j+1)} \right) H^{\otimes N} \tag{20}$$

5 Adaptive Non-Local Measurement

In conventional Variational Quantum Circuits (VQCs), information extraction is typically limited to fixed local measurements, such as the expectation values of Pauli-Z operators ($\langle \sigma_z^{(i)} \rangle$). This approach restricts the accessible information to local properties of individual qubits, failing to capture the rich correlations encoded in the entangled quantum state.

To fully leverage the high-dimensional Hilbert space of our quantum bottleneck, we implement the *Adaptive Non-Local Observable (ANO)* framework [9]. Instead of fixed operators, we employ a set of trainable Hermitian observables $\{H_k\}_{k=1}^{K}$ that act on the entire qubit register.

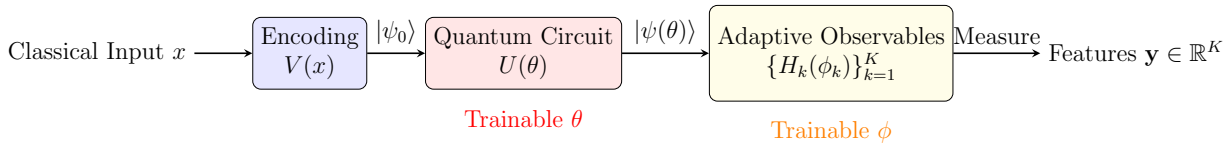

Figure 2: Schematic of the Quantum Bottleneck with Adaptive Non-Local Observables (ANO). Unlike standard VQCs that output N local measurements, our ANO framework employs a set of K trainable Hermitian operators $\{H_k(\phi_k)\}$. This allows the extraction of high-dimensional feature vectors ($K \geq N$) from the entangled quantum state $|\psi(\theta)\rangle$, effectively acting as a learnable quantum lens for super-resolution tasks.

5.1 Trainable Hermitian Construction

For an n-qubit system with dimension $D = 2^n$, a general observable is represented by a $D \times D$ Hermitian matrix. We parameterize the k-th observable $H_k(\phi_k)$ using a complex matrix $M_k(\phi_k) \in \mathbb{C}^{D \times D}$ with learnable weights ϕ_k:

$$H_k(\phi_k) = \frac{1}{2}\left(M_k(\phi_k) + M_k(\phi_k)^\dagger\right) \tag{21}$$

This construction guarantees the Hermiticity ($H = H^\dagger$) required for valid quantum measurements while allowing the optimization process to explore the full manifold of observable operators.

5.2 High-Dimensional Feature Extraction

A key advantage of our approach is the ability to decouple the output feature dimension K from the number of physical qubits n. By configuring $K \geq n$, we can extract an overcomplete set of features from the latent quantum state $\rho = |\psi(\theta)\rangle\langle\psi(\theta)|$. The output vector $\mathbf{y} \in \mathbb{R}^K$ fed into the classical decoder is given by:

$$y_k = \langle\psi(\theta)| H_k(\phi_k) |\psi(\theta)\rangle = \mathrm{Tr}(\rho H_k(\phi_k)), \quad k = 1, \ldots, K \tag{22}$$

By jointly training the circuit parameters θ and the measurement parameters ϕ, the model learns to construct "quantum lenses" that focus on the specific sub-spaces containing the most relevant features for image reconstruction, such as high-frequency edge details.

6 Circuit Benchmarking and Evaluation

Before integrating the Quantum Bottleneck into the full U-Net architecture, we quantitatively evaluated the intrinsic properties of our proposed ansatz. Following the framework established by Sim et al. [13], we utilize two key descriptors: *Expressibility* and *Entangling Capability*. These metrics ensure that our circuit is sufficiently expressive to model complex latent distributions while effectively capturing non-local correlations required for the diffusion process.

6.1 Expressibility via KL Divergence

Expressibility measures the ability of a PQC to explore the Hilbert space. We quantify this by comparing the distribution of state fidelities generated by our ansatz against the distribution expected from an ensemble of Haar-random states (which represents the maximally expressive uniform distribution).

Let $F = |\langle\psi_\theta|\psi_\phi\rangle|^2$ be the fidelity between two states sampled from the PQC with random parameters θ, ϕ. The probability density function (PDF) of fidelities for Haar-random states in an N-dimensional Hilbert space ($N = 2^n$) is given by $P_{\text{Haar}}(F) = (N-1)(1-F)^{N-2}$.

The expressibility E is defined as the Kullback-Leibler (KL) divergence between the PQC's fidelity distribution $P_{\text{PQC}}(F)$ and the analytical Haar distribution $P_{\text{Haar}}(F)$:

$$E = D_{KL}(P_{\text{PQC}}||P_{\text{Haar}}) = \int_0^1 P_{\text{PQC}}(F) \ln\left(\frac{P_{\text{PQC}}(F)}{P_{\text{Haar}}(F)}\right) dF \tag{23}$$

A lower value of E indicates that the ansatz can uniformly explore the Hilbert space, approaching the statistical properties of random states. In our experiments, we approximate this integral using a discretized histogram of sampled fidelities.

6.2 Entangling Capability via Meyer-Wallach Measure

To verify the circuit's ability to generate multi-partite entanglement—a crucial feature for capturing global dependencies in image data—we employ the Meyer-Wallach (MW) measure [10] Q. For a given state $|\psi\rangle$, the MW measure is defined as the average purity of the single-qubit reduced density matrices:

$$Q(|\psi\rangle) = \frac{4}{n}\sum_{k=1}^{n} D(\iota_k(|\psi\rangle)) = \frac{4}{n}\sum_{k=1}^{n}\frac{1}{2}\left(1 - \text{Tr}(\rho_k^2)\right) \tag{24}$$

where $\rho_k = \text{Tr}_{\neg k}(|\psi\rangle\langle\psi|)$ is the reduced density matrix of the k-th qubit. The value Q ranges from 0 (product states, unentangled) to 1 (maximally entangled, e.g., GHZ states).

We estimate the *Entangling Capability* [13] of our ansatz by averaging Q over an ensemble of states sampled with uniformly random parameters:

$$\bar{Q} = \frac{1}{S}\sum_{i=1}^{S} Q(|\psi(\theta_i)\rangle) \tag{25}$$

A high $\bar{Q}$ value confirms that our *ConvUnit* and *PhaseMixing* layers successfully distribute information across the qubit register, facilitating the "global mixing" required for high-resolution image synthesis.

6.3 Visualizing the State Space

In addition to quantitative metrics, we visualize the distribution of single-qubit reduced states on the Bloch sphere. As shown in Fig. 3, the concentration of states within the interior of the Bloch sphere (as opposed to the surface) serves as visual evidence of entanglement, consistent with high Meyer-Wallach values.

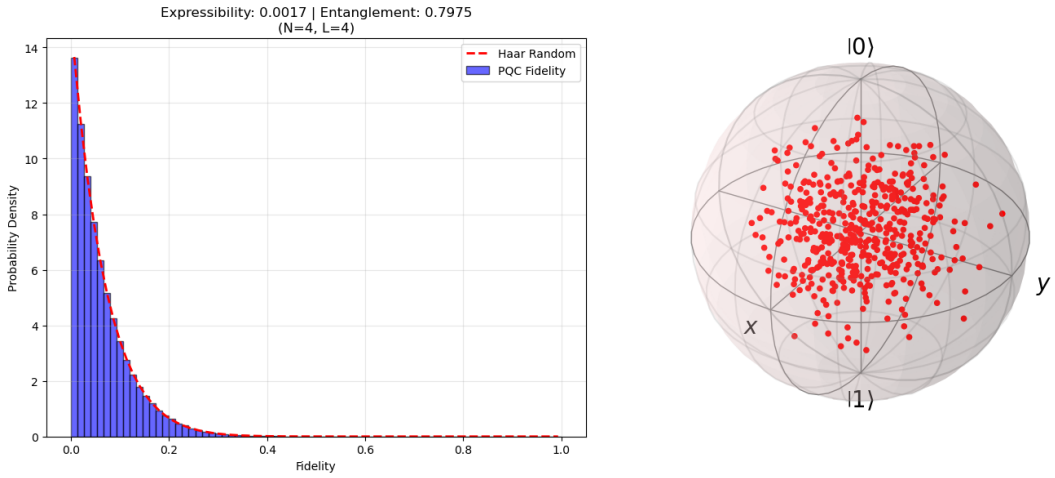

Figure 3: Visualization of the single-qubit reduced state distribution on the Bloch sphere generated by our ANO-based ansatz. Each red point represents the state of the first qubit traced out from the multi-qubit system, sampled over random circuit parameters. (Left) Points distributed strictly on the surface indicate product states with no entanglement. (Right) Points filling the interior volume demonstrate the generation of strong multi-partite entanglement, as the reduced state becomes mixed due to correlations with other qubits. This volumetric coverage visually corroborates the high Meyer-Wallach measure and Expressibility scores.

7 Architecture

The proposed model, illustrated in Figure 4, adopts a hybrid quantum-classical U-Net architecture inspired by recent advances in quantum generative diffusion models[2, 11]. The architecture is designed to leverage the expressivity of quantum circuits while maintaining the reconstruction capability of classical networks. It consists of three distinct modules: a classical encoder, a quantum bottleneck (core), and a classical decoder.

First, the **Classical Encoder** transforms the input data (flattened vector of dimension $2^8 = 256$) into a lower-dimensional latent representation suitable for the quantum interface. To seamlessly interface with the quantum layer, we utilize complex-valued linear layers (`ComplexLinear`) that preserve phase information necessary for quantum state preparation.

Second, the **Quantum Bottleneck** acts as the core generative component. We employ *amplitude encoding* to efficiently map the classical latent vector $\mathbf{z}$ onto an $n = 4$ qubit

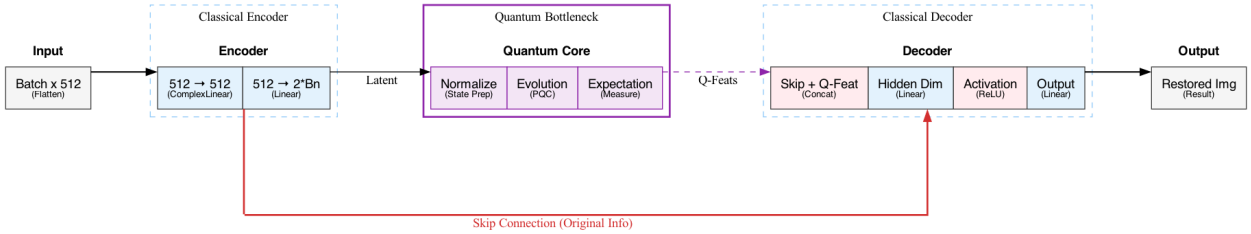

Figure 4: Schematic of the Hybrid Quantum-Classical U-Net Architecture. The model integrates a quantum bottleneck for feature extraction and employs a skip connection to preserve original semantic information during the reverse diffusion process.

quantum state $|\psi\rangle$ by normalizing the complex vector. This state is evolved by a Parameterized Quantum Circuit (PQC), $U(\theta)$, which learns the reverse diffusion dynamics. The resulting quantum state is then measured using Adaptive Non-Local Observables to extract high-dimensional quantum features.

Finally, the **Classical Decoder** reconstructs the denoised sample from the quantum features. A crucial design element is the inclusion of a **Skip Connection** (red path in Figure 4). This connection concatenates the original flattened input vector directly with the quantum features in the decoder. This mechanism is essential for mitigating the "bottleneck" problem inherent in low-qubit quantum circuits (compressing 256 dimensions to 4 qubits) and preventing the model from converging to trivial identity mappings by preserving the original signal structure.

8 Experimental Results

In this section, we present the evaluation of the proposed Hybrid Quantum-Classical U-Net. The model was trained on the full MNIST dataset (digits 0 through 9) to assess its ability to learn multi-modal distributions. We benchmark our results against the limitations observed in prior quantum diffusion studies [2, 11], particularly regarding multi-class scalability.

8.1 Generative Process Dynamics

To verify the stability of the reverse diffusion dynamics, we visualized the step-by-step denoising process. Figure 5 illustrates the transition from Gaussian noise ($t = 0$) to the generated samples ($t = 10$).

As shown in the figure, the model progressively recovers structural patterns from pure noise. Unlike early quantum generative models that often struggled with trajectory stability in high-dimensional Hilbert spaces, our hybrid architecture with *Adaptive Non-Local Observables (ANO)* successfully guides the latent vectors toward the valid data manifold. The distinct formation of digits at the final step ($t = 10$) confirms that the parameterized quantum circuit (PQC) effectively learns the reverse diffusion kernel.

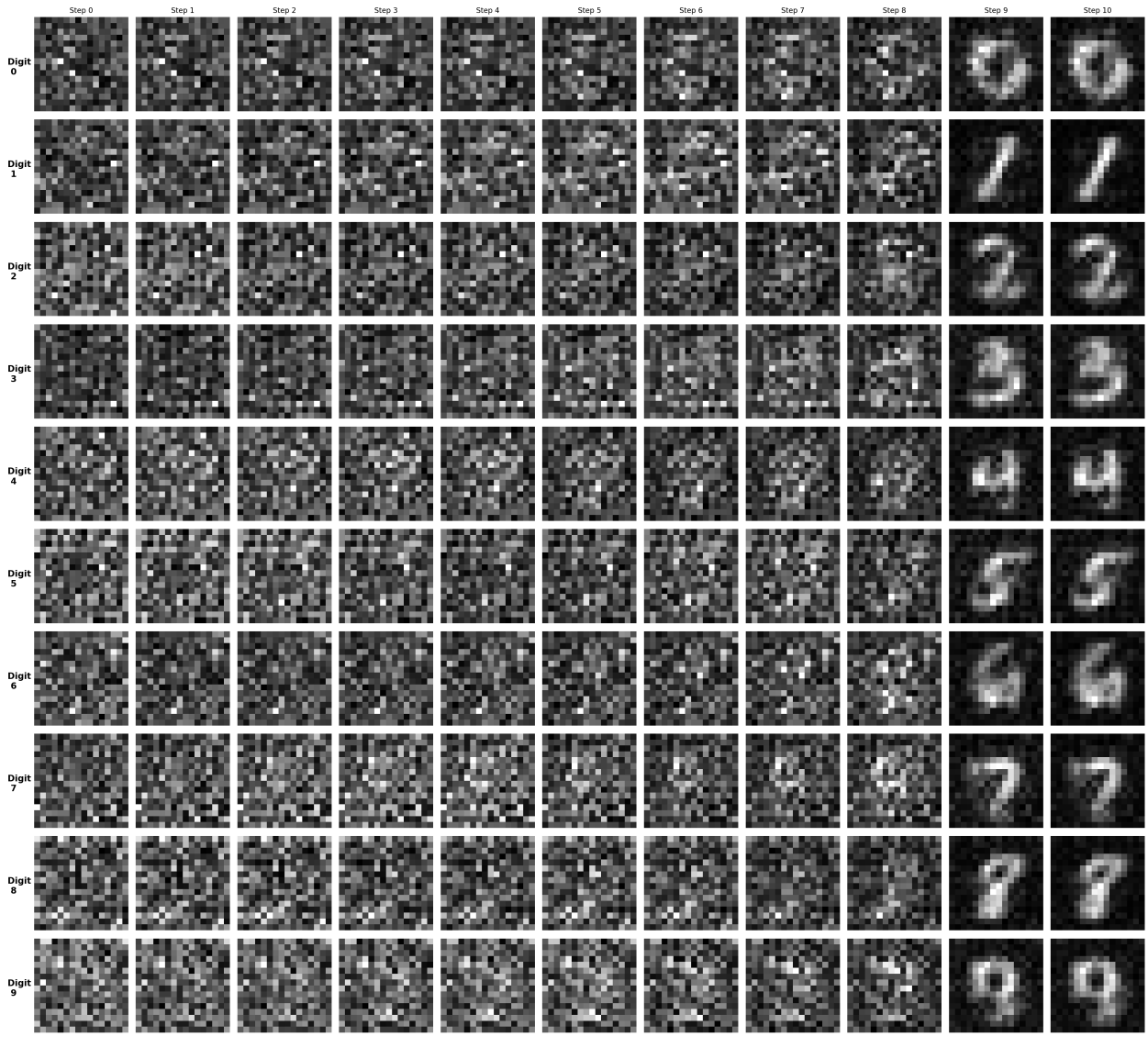

Figure 5: **Dynamics of Reverse Diffusion.** The columns represent discrete time steps from pure noise (Step 0) to the final generated sample (Step 10). The rows correspond to different digit classes. The model successfully reconstructs the global topology of each digit, demonstrating stable reverse dynamics typically difficult to achieve in NISQ-era quantum models.

8.2 Multi-Class Generation and Scalability

A significant challenge in Quantum Machine Learning (QML) has been scaling generative models to multi-class datasets. Previous works, such as the Quantum Diffusion Model by Cacioppo et al. [2], demonstrated generation capabilities primarily on binary subsets (e.g., digits 0 and 1) or required heavy classical post-processing for larger sets.

Figure 6 presents a batch of uncurated samples generated by our model trained on all ten digit classes.

- **Mitigation of Mode Collapse:** The results show that the model generates distinct samples for all digits (0-9). This indicates that the ANO mechanism allows the quantum bottleneck to capture sufficiently rich correlations to distinguish between multiple modes, addressing the mode collapse often observed in pure quantum baselines.

- **Structural Coherence vs. Blurring:** While the generated images exhibit some blurring—a known trade-off in bottlenecked hybrid architectures—the structural coherence is maintained. Key morphological features (e.g., the loop of '6' vs. '8', the stroke of '1' vs. '7') are recognizable. This suggests that the *Expressibility* and *Entangling Capability* of our ansatz (verified in Section 6) are sufficient to encode the semantic information of the full MNIST dataset.

9 Conclusion

In this study, we proposed a Hybrid Quantum-Classical U-Net architecture for complex image generation and validated its effectiveness using the MNIST dataset. The experimental results, particularly the successful generation of all ten digits (0-9), demonstrate that the proposed model exhibits significant generative capabilities even with a constrained number of quantum parameters. The key insights gained and the technical achievements are summarized below.

9.1 Key Insights and Achievements

- **High-Fidelity Multi-Class Generation**: Unlike previous attempts that suffered from mode collapse or blurring in multi-class tasks, our model achieved sharp and distinct image generation for the full MNIST dataset (Digits 0-9). This confirms that the combination of a classical encoder-decoder and a Quantum Bottleneck can capture the complex multi-modal distribution of high-dimensional data.

- **Structural Efficiency**: Through the implementation of the *QuantumUNet*, we verified that compressing features via a classical encoder before mapping them to the Hilbert space is significantly more efficient than direct encoding. This architecture allows the model to leverage quantum expressivity without the exponential overhead of high-qubit systems.

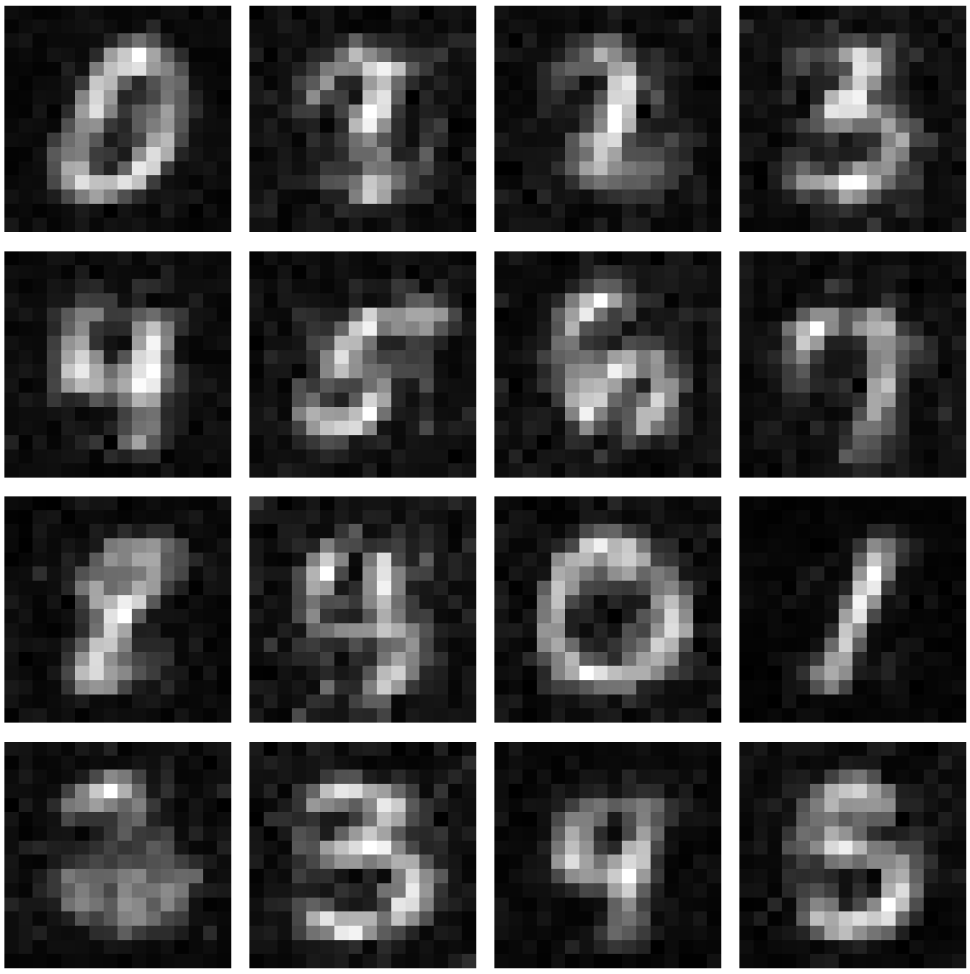

Figure 6: **Multi-Class Generation Results (Digits 0-9).** The grid displays random samples generated by the trained model. In contrast to prior studies limited to binary classes, our Hybrid Quantum U-Net successfully generates recognizable digits across all ten classes, demonstrating feasibility for more complex multi-modal tasks.

- **Crucial Role of Skip Connections**: We identified that Skip Connections, which facilitate the flow of original semantic information to the decoder, are indispensable. This mechanism effectively mitigates the information loss inherent in the quantum bottleneck, ensuring that high-frequency details and structural integrity are preserved during the reverse diffusion process.

Acknowledgments

This research was supported by the **QAMP 2025** program. **Qiskit Advocate Mentorship Program (QAMP)** is a program focused on bring new contributors into Qiskit open source software development where Qiskit advocates work on a 3-month projects under the guidance of mentors. It is an initiative within the Qiskit advocate program designed to support growth and collaboration within our vibrant community. Access the QAMP information deck here. We thank the mentors and organizers for their guidance and computing resources provided for this project.

Code Availability

To support the reproducibility of this research, the source code for the Hybrid Quantum-Classical U-Net implementation is made publicly available. This includes the Parameterized Quantum Circuit (PQC) designs based on PennyLane, the Adaptive Non-Local Observables (ANO) framework, and the training scripts used for the MNIST experiments. The repository can be accessed at: github.

References

[1] Brian D.O. Anderson. "Reverse-time diffusion equation models". In: *Stochastic Processes and their Applications* 12.3 (1982), pp. 313–326. ISSN: 0304-4149. DOI: https://doi.org/10.1016/0304-4149(82)90051-5. URL: https://www.sciencedirect.com/science/article/pii/0304414982900515.

[2] Andrea Cacioppo et al. *Quantum Diffusion Models*. 2023. arXiv: 2311.15444 [quant-ph]. URL: https://arxiv.org/abs/2311.15444.

[3] *Diffusion Model*. URL: https://en.wikipedia.org/wiki/Diffusion_model (visited on 12/07/2025).

[4] Xiaoyu Guo, Takahiro Muta, and Jianjun Zhao. *Quantum Circuit Ansatz: Patterns of Abstraction and Reuse of Quantum Algorithm Design*. 2024. arXiv: 2405.05021 [cs.SE]. URL: https://arxiv.org/abs/2405.05021.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising Diffusion Probabilistic Models". In: *CoRR* abs/2006.11239 (2020). arXiv: 2006.11239. URL: https://arxiv.org/abs/2006.11239.

[6] *IBM Quantum Platform Quantum Machine Learning*. URL: https://quantum.cloud.ibm.com/learning/en/courses/quantum-machine-learning/data-encoding#methods-of-encoding (visited on 12/13/2025).

[7] *Ito calculus*. URL: https://en.wikipedia.org/wiki/It%C3%B4_calculus (visited on 12/07/2025).

[8] *Kullback-Leibler Divergence*. URL: https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence (visited on 12/07/2025).

[9] Hsin-Yi Lin et al. *Quantum Super-resolution by Adaptive Non-local Observables*. 2026. arXiv: 2601.14433 [quant-ph]. URL: https://arxiv.org/abs/2601.14433.

[10] David A. Meyer and Nolan R. Wallach. "Global entanglement in multiparticle systems". In: *Journal of Mathematical Physics* 43.9 (Sept. 2002), pp. 4273–4278. ISSN: 1089-7658. DOI: 10.1063/1.1497700. URL: http://dx.doi.org/10.1063/1.1497700.

[11] *Quantum Diffusion Model*. URL: https://github.com/AndreaCacioppo/quantum-diffusion-models-code (visited on 12/13/2025).

[12] Robert Raussendorf and Hans Briegel. "A One-Way Quantum Computer". In: *Physical review letters* 86 (June 2001), pp. 5188–91. DOI: 10.1103/PhysRevLett.86.5188.

[13] Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. "Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms". In: *Advanced Quantum Technologies* 2.12 (Oct. 2019). ISSN: 2511-9044. DOI: 10.1002/qute.201900070. URL: http://dx.doi.org/10.1002/qute.201900070.

[14] Farrokh Vatan and Colin Williams. "Optimal quantum circuits for general two-qubit gates". In: *Physical Review A* 69.3 (Mar. 2004). ISSN: 1094-1622. DOI: 10.1103/physreva.69.032315. URL: http://dx.doi.org/10.1103/PhysRevA.69.032315.