Denoising Diffusion Models with Improved Quantum Implicit Neural Networks for Image Generation

ABSTRACT

Quantum generative models leverage principles of quantum mechanics to address the rising computational demands of classical generative methods. Among them, quantum denoising diffusion models (QDDMs) offer improved training stability and image quality. However, current QDDMs face two key challenges: (1) limited capacity to learn complex data distributions, reducing image fidelity and diversity; and (2) reliance on full-step sampling and complex UNet-like quantum architectures, hindering deployment on near-term devices. To address these, we propose the Quantum Implicit Denoising Diffusion Models (QIDDM), which employs the improved quantum implicit neural network to boost expressiveness. Two efficient variants—a hybrid and a fully quantum model—are introduced, incorporating consistency models to enable $O(1)$ image sampling and improve compatibility with near-term devices. Extensive experiments on diverse datasets demonstrate the superiority of QIDDM. Noise analysis and validation across three superconducting quantum computers further confirm its practical potential in advancing quantum generative modeling. Code is available at: https://github.com/SC2025anonymous288/QIDDM.

1 INTRODUCTION

Image generation remains a key area of research in computer vision and graphics, encompassing a wide range of applications from synthetic data creation to artistic endeavors. Generative models, such as Stable Diffusion, have found diverse applications, including image editing and supporting multi-modal models like GPT-4 [1] in generating visual responses to human queries. These models have made significant technological advancements and have had a profound impact on our daily lives [34]. At their core, these models aim to learn the underlying probability distributions of a given dataset. Once trained, they can generate new data that follows the same distribution, thereby demonstrating an understanding of the inherent patterns in the data.

Quantum generative models [42]—the quantum analogs of classical generative methods in quantum machine learning [7, 36, 38]—have attracted considerable interest due to their promise of leveraging quantum mechanics for data synthesis. Prominent examples include Quantum Generative Adversarial Networks (QGANs) [26, 43], Quantum Circuit Born Machines (QCBMs) [5, 24], Quantum Variational Autoencoders (QVAEs) [18], and Quantum Boltzmann Machines (QBMs) [2]. Among these, QGANs, which extend classical GANs [9, 12, 15, 23] to the quantum domain, have been explored for quantum state generation, image synthesis, and classical distribution fitting. However, QGANs inherit many of the same challenges as their classical counterparts—particularly instability and mode collapse, which cause oscillations in generator–discriminator losses and impede convergence to a global optimum [8, 10, 30].

On the other hand, researchers have developed **quantum denoising diffusion models (QDDMs)** to address the limitations of QGANs. Some studies [11, 50] use trace-out measurements to place hybrid quantum-classical subsystems into mixed states and train quantum neural networks (QNNs) [4] with parameterized quantum circuits (PQCs) [6], enabling subsystem states to approximate target images. However, manipulating and training mixed-state systems remains challenging due to quantum hardware limitations, and these methods require separate PQC parameter updates at each time step, making them unsuitable for Noisy Intermediate-Scale Quantum (NISQ) devices [32]. Recent work [20] has introduced QDDMs with **Consistency Models** [41], normalizing training into a unitary single-sampling framework, reducing training time and improving NISQ compatibility. However, the Qdense model proposed in this study relies on amplitude encoding, which is qubit-efficient but difficult to prepare and struggles to capture crucial image information [4, 6]. Additionally, a hybrid quantum diffusion model based on a quantum U-net framework [14, 20] modifies convolutional layers into quantum ones and integrates modules like ResNet [47] and Attention [3]. While these adaptations enable hybrid modeling, they result in slow training speeds, and frequent quantum-classical data conversions introduce additional errors and limit the model's applicability. Moreover, existing quantum denoising diffusion models suffer from a lack of theoretical interpretability and have rarely been applied to image generation tasks, particularly in complex domains like facial image generation [31].

To address these challenges, we present the **Quantum Implicit Denoising Diffusion Model (QIDDM)**, which draws inspiration from Quantum Implicit Neural Networks (QINNs) [51] and the Consistency Model [41]. The key insight is that, by using our proposed improved QINNs (IQINNs) to learn a continuous reverse-diffusion mapping, QIDDM can achieve $O(1)$ single-step sampling—eliminating the need to train separate PQCs at each diffusion step. We offer two variants to balance expressiveness with hardware feasibility:

- **QIDDML (Hybrid):** a lightweight model that integrates a single classical linear layer after quantumness, preserving a shallow circuit while still outperforming angle-encoded QNNs with fewer parameters.
- **QIDDMA (Fully Quantum):** a purely quantum architecture that directly extracts amplitude information, sidestepping classical post-processing and surpassing QDense in image quality under the same parameter budget.

Both designs greatly simplify the hybrid QINN framework yet retain its capacity to model complex data distributions. To demonstrate real-world feasibility, we benchmark QIDDM on 28×28 MNIST [21], Fashion-MNIST [46], and EMNIST [13], comparing against state-of-the-art QDense [20], baseline QNNs, and classical U-Net-based denoising diffusion models [35]. Our results show that QIDDM variants consistently outperform these baselines in terms of both parameter efficiency and image quality. We also compare against the latest quantum GAN (PQWGAN [43]) and classical GAN (WGAN-GP [9]) under equivalent conditions: even using only 1% of their parameter counts, our models deliver superior image fidelity.

Furthermore, we extend the application of QIDDM to facial image and medical generation, using CelebA (64×64) and MEDICAL. To our knowledge, this is the first detailed demonstration of generating complex images with a quantum diffusion framework, achieving impressive results with a remarkably low number of network parameters. Finally, we study the impact of different quantum noise types on model performance and conduct experiments on three superconducting quantum computers to validate noise robustness and hardware compatibility.

In summary, our contributions are:

- **QIDDM Framework:** We propose a unified denoising diffusion approach that fuses improved QINNs (IQINNs) with a Consistency Model. This design enables a single IQINN to learn the entire reverse-diffusion process, reducing both parameter count and sampling complexity to $O(1)$ via unitary operations.
- **Hybrid and Fully Quantum Variants:** We introduce QIDDML (hybrid) and QIDDMA (fully quantum), each optimized for NISQ devices. QIDDML uses only one linear layer yet outperforms angle-encoded QNNs, while QIDDMA matches or exceeds QDense in quality with the same parameter scale.
- **Comprehensive Evaluation and Real-Device Validation:** We demonstrate that our models achieve state-of-the-art image quality with less than 1% of the parameter count of QGANs and classical GANs. We also present the application of QIDDM to complex image generation, and validate model robustness through quantum noise analysis and real-device experiments on three superconducting quantum platforms.

2 PRELIMINARY AND RELATED WORK

2.1 Diffusion Models

Diffusion models, introduced by Sohl-Dickstein et al. [39], have emerged as a powerful class of generative models. Early implementations demonstrated promise, and subsequent work—such as Denoising Diffusion Implicit Models (DDIM) [40]—has further improved sampling efficiency by removing noise early and skipping certain Markov steps [17, 28].

Given a training dataset $\mathcal{D}$ and data $\mathbf{x} \in \mathcal{D}$, a diffusion model gradually corrupts $\mathbf{x}_t$ by adding noise over T steps. Concretely, the forward diffusion process is:

$$q(\mathbf{x}_{0:T}) = q(\mathbf{x}_0) \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{1}$$

where $q(\mathbf{x}_0)$ represents the distribution of the clean data. In classical diffusion, one typically performs iterative denoising: starting from pure noise $\mathbf{x}_T$, a sequence of inverse steps produces $\mathbf{x}_T, \mathbf{x}_{T-1}, \ldots, \mathbf{x}_0$. However, this iterative process can be both time-consuming and resource-intensive.

To address these drawbacks, *Consistency Models* (CMs) [41] learn a *single-step* reverse mapping. Specifically, CMs train a function $f_\theta(\mathbf{x}, t)$ that directly maps the noisy data $\mathbf{x}_t$ at any time t back to the clean images $\mathbf{x}_0$. Formally, the consistency function takes the form

$$f_\theta(\mathbf{x}, t) = c_{skip}(t)\mathbf{x} + c_{out}(t)F_\theta(\mathbf{x}, t), \tag{2}$$

where $c_{skip}(t)$ and $c_{skip}(t)$ are differentiable functions such that $c_{skip}(0) = 1$, and $c_{out}(0) = 0$. Here, $F_\theta(\mathbf{x}, t)$ is a neural network parameterized by θ.

The model is trained by minimizing the following loss function:

$$\arg\min_\theta \|f_\theta(\mathbf{x}_t) - \mathbf{x}_0\|, \tag{3}$$

where $\mathbf{x}_t$ is obtained via a scheduling function w_t that grows from 0 to 1 over $t = 0, \ldots, T$. At inference time, we simply apply

$$f_\theta(\mathbf{x}_T, T) = \mathbf{x}_0, \tag{4}$$

where $\mathbf{x}_T$ is the known Gaussian distribution $\mathcal{N} \sim (\mu_T, \sigma_T^2)$ at T time step in forward process. This single-step mapping greatly reduces sampling complexity compared to iterative methods.

2.2 Quantum Neural Networks

Quantum machine learning (QML) [7] seeks to exploit quantum hardware to tackle tasks that are challenging for classical algorithms [42]. At the heart of many QML approaches lie *parameterized quantum circuits* (PQCs), which act as trainable function approximators analogous to classical neural networks.

One influential idea in quantum architectures is *data re-uploading*. Perez-Salinas et al. [33] showed that by interleaving data-encoding rotations with trainable gates, a single-qubit circuit can approximate arbitrary functions. Subsequent works [37, 49] proved that data re-uploading circuits are mathematically equivalent to truncated Fourier series, thus providing universal approximation capabilities. Building on this, [51] introduced *Quantum Implicit Neural Networks* (QINNs), which use multi-qubit data re-uploading modules plus entangling layers to capture high-frequency information and achieve richer expressivity.

2.3 Quantum Diffusion Models

Several recent works have extended classical diffusion to the quantum domain. Early attempts [50] construct a dedicated PQC for each diffusion time step: at each step, the model measures a subset of qubits (trace-out) and feeds the results into a classical controller, forming a hybrid quantum–classical system. While conceptually appealing, this approach suffers from exploding parameter counts and requires separate PQC parameter updates at each time step, making it impractical on NISQ devices.

To mitigate parameter growth, Chen et al. [11] propose a *parameter-sharing* strategy: the same PQC is reused across all time steps, reducing the total number of trainable parameters. Nevertheless, these methods still rely on mixed-state manipulations and frequent quantum–classical feedback, which incur substantial overhead and noise accumulation.

An alternative line of research explores hybrid Unet-based architectures. For instance, Reference [14] proposes two hybrid Unet variants in which classical convolutional and ResNet modules are replaced by quantum counterparts. The most recent advancement, Reference [20], integrates a consistency model tailored to the noise resilience of quantum hardware. However, both models rely on a hybrid Unet structure featuring quantum convolutional neural networks and QDense—a deep quantum neural network based on amplitude encoding. Due to the no-cloning theorem [7], the skip connections in the Unet architecture cannot be directly implemented

Table 1: Comparison of Recent Quantum Diffusion models and our proposal QIDDM.

Domain	Denoising Diffusion Models (DDMs)			
Methodology	Diffusion Principle			
Models	QuDDPM [50]/QGDM [11]	Qdense [20]/QNN	Hybrid U-net [14]	QIDDM (Ours)
Core Modules	Strongly entangling quantum circuit / Mixed state system	Strongly entangling quantum circuit	Quantum convolutional neural network, Resnet, attention module	Quantum Implicit Neural Network
Results	More parameters, long training time, NISQ-unfriendly	Low quality generated images / More parameters	Complex, very long training time and difficult to implement	Fewer parameters, more NISQ-friendly, high-quality image generation

on quantum hardware. Moreover, the inherent complexity of Unet, which involves convolutional layers, pooling layers, ResNet blocks, and attention mechanisms, necessitates frequent quantum–classical data exchanges. These exchanges introduce significant overhead and are prone to elevated error rates on current hardware.

Overall, existing quantum diffusion approaches either incur excessive resource requirements or lack a clear theoretical interpretation—especially for complex image synthesis tasks such as face generation [31]. To address this gap, we propose the Quantum Implicit Denoising Diffusion Model (QIDDM) framework, which combines the improved quantum implicit neural networks with consistency model. We present two variants—QIDDML (hybrid) and QIDDMA (fully quantum)—and compare them with existing methods in Table 1.

3 QUANTUM IMPLICIT DENOISING DIFFUSION MODELS

3.1 Forward Diffusion Process

To make our model NISQ-compatible, we perform the entire forward (noise-adding) process classically, thereby conserving quantum resources. As illustrated in Figure 1(a), given a dataset $\mathcal{D}$, each single clean image $x \in \mathcal{D}$ is corrupted over T steps until it approximates a Gaussian noise distribution ϵ_T.

Concretely, at time step t, we generate a noisy image x_t by linearly interpolating between the original image x_0 and independent Gaussian noise $\epsilon_t \sim \mathcal{N}(\mu_T, \sigma_{T^2})$:

$$x_t = (1 - w_t) \cdot x_0 + w_t \cdot \epsilon_t \tag{5}$$

where $w_t \in [0, 1]$ is a noise weight that increases with t. In the initial step ($t = 0$), $w_0 = 0$, so x_0 remains noise-free; in the final step ($t = T$), $w_T = 1$, so $x_T \approx \epsilon_T$. As t increases, x_t gradually transitions from the data distribution $p_{data}(x)$ to the noise distribution $\epsilon_t \sim \mathcal{N}(\mu_T, \sigma_{T^2})$.

We set w_t using a simple power schedule over a linearly spaced grid:

$$w_t = \frac{\text{linspace}(0, 1, T)^q}{\max(\text{linspace}(0, 1, T)^q)}, \tag{6}$$

where linear space $(0, 1, T)$ generates T evenly spaced values in $[0, 1]$, and $q > 0$ controls the rate at which noise is added. Because this entire forward process is classical, we incur no quantum circuit overhead until the backward (denoising) stage. In other words, quantum resources are reserved exclusively for learning the reverse mapping $x_T \to x_0$, allowing us to focus on a shallower, more noise-resilient quantum circuit.

3.2 Backward Denoising Process

The backward denoising process is executed on a quantum computer. Leveraging the consistency model, we train a quantum implicit neural network (IQINN) to map any noisy input x_t directly to its clean counterpart $\hat{x}_0$ in a single step. As depicted in Figure 1(b), the trained model QIDDM and its consistency function $\hat{f}_\theta$, parameterized by θ, should satisfy

$$\hat{f}_\theta(x_t, t) = \hat{x}_0 \approx f(x_t, t) = x_0, \tag{7}$$

where f denotes the ideal consistency function (see Equations (2) and (4)). In other words, $\hat{f}_\theta(x_t, t)$ approximates the true reverse-diffusion mapping from the noisy sample x_t at time step t back to the original image x_0.

To obtain $\hat{f}_\theta$, we note that the IQINN is trained to predict the noise term ϵ_t added during the forward diffusion. Concretely, for each $t \in \{1, \ldots, T\}$, the network output

$$\hat{F}_\theta(x_t) = \hat{\epsilon}_t \approx \epsilon_t. \tag{8}$$

Here, $\hat{F}_\theta(x_t)$ denotes the IQINN's estimate of the true noise ϵ_t at step t. Assuming a simplified forward process with linear noise injection (Equation (5)), the consistency function can be expressed as

$$\hat{f}_\theta(x_t, t) = \frac{1}{1 - w_t} x_t - \frac{w_t}{1 - w_t} \hat{F}_\theta(x_t), \tag{9}$$

where w_t is defined in Equation 6. Substituting $\hat{F}_\theta(x_t) \approx \epsilon_t$ into Equation 9 yields $\hat{f}_\theta(x_t, t) \approx x_0$, thus achieving one-step denoising.

In summary, as illustrated in Figure 1(c), we train our improved quantum implicit neural network $\hat{F}_\theta$ to estimate the noise ϵ_t added at each time step t, thereby ensuring that the consistency function $\hat{f}_\theta$ closely approximates the ideal reverse-diffusion mapping f.

3.3 Quantum Circuit Construction of QIDDM

As shown in Figure 2, IQINNs are an improved variant of QINNs that preserve the data re-uploading architecture. To reduce the overhead and error accumulation from repeated classical–quantum data transfers in an N-layer cycle—and to better match NISQ-era constraints—we introduce two streamlined variants, QIDDML and QIDDMA.

Classical Components: QIDDML requires only:

- A single linear layer to expand the n-dimensional quantum measurement vector back to the original d^2 dimensions.
- A classical optimizer (e.g. Adam [19]) for gradient-based training.

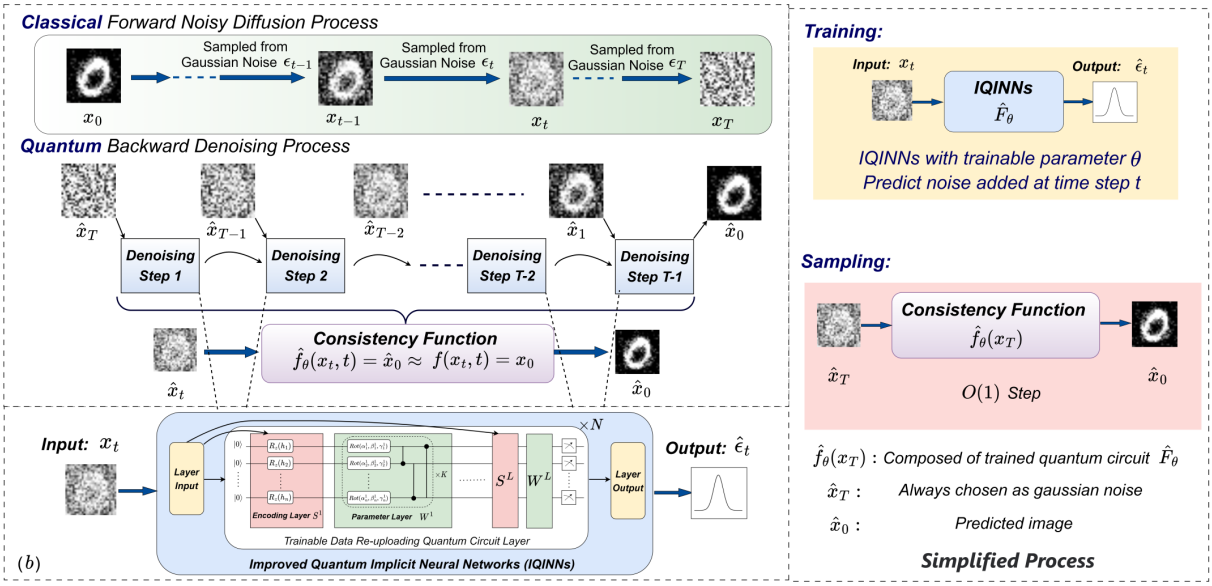


Figure 1: Architecture of QIDDM. (a) presents the overall workflow of QIDDM. (b) demonstrates the detailed quantum circuit construction based on Quantum Implicit Neural Networks. (c) shows the simplified process of our proposed QIDDM.

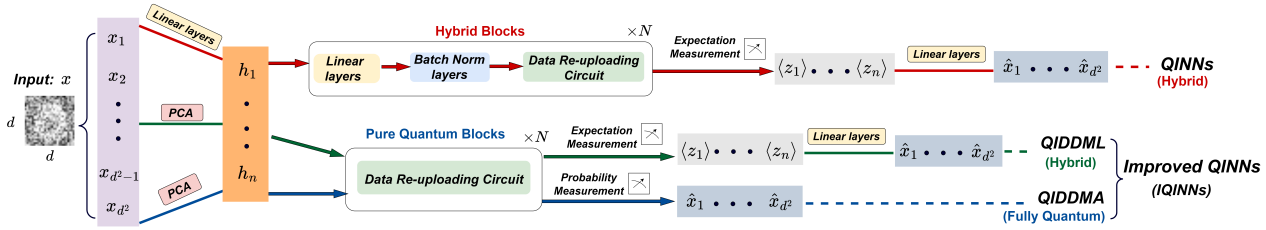


Figure 2: Comparison of QINNs and our proposal Improved QINNs with two models QIDDML and QIDDMA.

By contrast, the original QINNs model (Figure 2 red line workflow) uses multiple linear layers, batch-normalization, and repeated data re-uploading to reduce and then re-expand dimensionality [51], greatly increasing parameter count and implementation complexity. In QIDDML, we first apply PCA to map the input image $x \in \mathbb{R}^{d\times d}$ into an $n = \log_2(d^2)$-dimensional feature vector h. After the quantum circuit produces n expectation values $\langle z_1 \rangle, \ldots \langle z_n \rangle$, a single linear layer reconstructs the d^2-dimensional output, which is then used to compute loss and update parameters.

While QIDDMA represents a fully quantum framework: all trainable parameters are embedded directly within the quantum circuit, and—aside from the optimizer—there are no classical components. This architecture thus constitutes the most "pure" quantum neural network model.

Quantum Components: As show in Figure 1 (b), we use an n-qubit register initialized in $|0\rangle^{\otimes n}$ and apply N data-reuploading layers. Each layer $j = 1, \ldots, N$ implements the unitary of L **Encoding Layers** S and **Parameterized Layers** W as follows:

$$U(h^j) \;=\; \prod_{l=1}^{L} S^l(h^j) \cdot W^l, \tag{10}$$

where

$$S(h^j) \;=\; R_z(h_1^j) \otimes R_z(h_2^j) \otimes \cdots \otimes R_z(h_n^j)$$

denotes the angle encoding of input $h^j = (h_1^j, \ldots, h_n^j)$, and

$$W^l \;=\; \prod_{k=1}^{K} \Big[\bigotimes_{i=1}^{n} Rot(\alpha_i^k, \beta_i^k, \gamma_i^k) \Big] U_E$$

is the trainable block of K parameterized single-qubit rotations

$$Rot(\alpha, \beta, \gamma) = \begin{bmatrix} e^{-i(\alpha+\gamma)/2}\cos(\frac{\beta}{2}) & e^{-i(\alpha-\gamma)/2}\sin(\frac{\beta}{2}) \\ -e^{i(\alpha-\gamma)/2}\sin(\frac{\beta}{2}) & e^{i(\alpha+\gamma)/2}\cos(\frac{\beta}{2}) \end{bmatrix}$$

followed by a strong entangling layer U_E of CNOT gates. The input of j-th layer: $h^j = h_1^j, h_2^j, \ldots, h_n^j$ is obtained from the $j-1$ layers expectation $\langle z^{j-1} \rangle = \langle z_1^{j-1} \rangle, \ldots \langle z_n^{j-1} \rangle$. For the first layer ($j = 1$), $h^1 = (h_1^1, \ldots, h_n^1)$ denotes the initial classical features extracted from the $d \times d$ input (via PCA or a linear layer), as shown in Figure 2.

After N layers, the final quantum state is

$$|IQs\rangle \;=\; \prod_{j=1}^{N} U(h^j)\,|0\rangle^{\otimes n}. \tag{11}$$

Given $|IQs\rangle$, we extract quantum data via two strategies:

- **QIDDML:** measure an n-qubit observable M to obtain $\hat{F}_M(\theta_1) = \langle IQs|M|IQs\rangle \in \mathbb{R}^n$, then apply a single linear layer

$$\hat{F}_\theta \;=\; W_{\theta_2}\,\hat{F}_M(\theta_1) \;+\; b_{\theta_2}$$

to recover the d^2 outputs, where θ_1 is denoted as quantum neural networks parameter composed of (α, β, γ) and θ_2 is denoted as the parameters in the linear layer.

- **QIDDMA:** directly read out the amplitudes $a_i = \langle i|IQs\rangle$ in the computational basis, yielding $\hat{F}'_\theta = \{\alpha_1, \ldots, \alpha_{d^2}\}$ without any classical post-processing layer.

By contrast, the original QINNs model (Figure 2) includes N batch normalization and $N+2$ linear layers [51], which substantially increases parameter count and error rates due to frequent classical–quantum data. Our QIDDML and QIDDMA architectures eliminate these overheads while retaining the expressive power of the quantum network. In what follows, we demonstrate that $\hat{F}_\theta$ serves as an excellent approximator for noise fitting in QIDDM.

Theorem 3.1. *[Univariate Approximation by QINNs] For any univariate square-integrable function $F : [-\pi, \pi] \to R$, and for all $\delta > 0$, there exists a QINNs $U_{\theta,N,L,K}(x)$ such that $|\psi_x\rangle = U_{\theta,N,L,K}(x)\,|0\rangle^{\otimes n}$ satisfies:*

$$\|\langle\psi_x|\, M\, |\psi_x\rangle - cF(x)\| \le \delta \tag{12}$$

for some normalizing constant c and observable M.

In literature [37, 49], it is shown that the single-qubit data re-uploading model can be represented by a truncated Fourier series expansion and is a universal approximator for any square-integrable univariate function. Literature [51] further extends the data re-uploading model to multiple qubits and proves that this universal quantum Fourier neural network can be represented by Quantum Implicit Neural Network (QINNs). In our proposed QIDDML, the quantum neural network is based on QINNs, where the linear layers and the batch normalization layers between the N quantum layers are removed as shown in Figure 2. This simplifies the classical data processing component while retain the property of Theorem 3.1 and also serves as a universal approximator for univariate square-integrable functions.

Corollary 3.2. *[Noise approximation properties of QIDDML] For any Gaussian noise distribution $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$ and any $\delta > 0$, there exists a QIDDML circuit output $\hat{F}_\theta$ such that*

$$\left\|\hat{F}_\theta - c\,\epsilon\right\| \le \delta$$

for suitable c, where $\hat{F}_\theta$ is obtained via measurement and a single linear layer.

Proof. For any Gaussian distribution $\epsilon = \mathcal{N}(\mu, \sigma^2)$, the probability density is:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{13}$$

and its integration:

$$\int_{-\infty}^{\infty} |p(x)|^2\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\sigma^2}}. \tag{14}$$

Since the result of Equation (14) is constant, Equation (13) is therefore square-integrable. Hence, by Theorem 3.1, OQINNs can effectively approximate the univariate noise function ϵ_t at any time step in the forward diffusion process, thereby verifying Corollary 3.2. It follows that, within the denoising diffusion framework, QIDDM likewise achieves effective approximation of ϵ_t, providing rigorous theoretical support for the proposed model.

Algorithm 1 Quantum Implicit Diffusion Model (QIDDM) Training

Require: Dataset $\mathcal{D}$, n qubits OQINNs, total time step of the diffusion process T, learning rate η, weighting function $w(t)$, parameter λ, stop condition $iters$;
Ensure: $n = \log d^2$, d^2 the dimension of single sample $x \in \mathcal{D}$;

Init:$\theta \sim \mathcal{U}(0, 2\pi)$, $iters = 0$;
Repeat
  Sample $X_0 \sim \mathcal{D}$, $t \sim \mathcal{U}(0, T)$, $\epsilon_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$
  Compute $x_t = (1 - w_t) \cdot x_0 + w_t \cdot \epsilon_t$
  $\mathcal{L}(\theta) = \mathbb{E}_{t\sim\mathcal{U}(0,T)}\mathbb{E}_{\epsilon_t\sim\mathcal{N}(\mu_t,\sigma_t^2)}\, \|\hat{\epsilon}_t - \epsilon_t\|^2 + \lambda \left\|\hat{f}_\theta(x_t, t) - x_0\right\|^2$
  $\theta \leftarrow \theta - \eta\nabla_\theta \mathcal{L}(\theta)$
  $iters = iters + 1$
Until $\mathcal{L}(\theta)$ converges or the number of iterations reaches the maximum

Algorithm 2 Quantum Implicit Diffusion Model (QIDDM) Generation

Require: The optimal parameters θ^*, trained IQINNs $\hat{F}_{\theta^*}$ and consistency function $\hat{f}_{\theta^*}$, denoising steps p, mean μ_T and variance σ_T^2, precision requirement δ;
Init:p=0, Matrix $U(\theta^*) = \hat{F}_{\theta^*}$, Vector $\vec{\epsilon_T}$ from Gaussian distribution $\mathcal{N}(\mu_T, \sigma_T^2)$, measurement metrics $\mathbf{d}$;
Repeat
  $\vec{\epsilon_p} = U(\theta^*)\vec{\epsilon_T}$
  $\vec{\epsilon_T} = \vec{\epsilon_p}$
  $p = p + 1$
Until $\mathbf{d}(\hat{f}_{\theta^*}(\vec{\epsilon_p}), x_0) \le \delta$ or the number of denoising steps reaches the maximum P

3.4 Training and Generation

Taking learning noise distribution ϵ_t as an example, the training objective of QIDDM is to map a noisy sample x_t at time step t to an estimate of the noise,

$$\hat{F}_\theta(t) = \hat{\epsilon}_t \approx \epsilon_t,$$

via the IQINN. The loss function of our QIDDM with CMs is defined as follows:

$$\mathcal{L} = \mathcal{L}_{denoising} + \lambda\mathcal{L}_{consistency} \tag{15}$$

$$= \mathbb{E}_{t\sim\mathcal{U}(0,T)}\mathbb{E}_{\epsilon_t\sim\mathcal{N}(\mu_t,\sigma_t^2)}\, \|\hat{\epsilon}_t - \epsilon_t\|^2 + \lambda \left\|\hat{f}_\theta(x_t, t) - x_0\right\|^2. \tag{16}$$

The term $\mathcal{L}_{\text{denoising}}$ represents the loss function for the prediction of noise by IQINNs, which measures the discrepancy between the predicted noise $\hat{\epsilon}_t$ and the true noise ϵ_t. $\mathcal{L}_{\text{consistency}}$ is denoted as the consistency loss. λ is a parameter used to balance the losses $\mathcal{L}_{denoising}$ and $\mathcal{L}_{consistency}$.

Training is performed as detailed in Algorithm 1, with an added **STOP** condition of a maximum iteration count $iters$ to prevent divergence. In each iteration, we compute $\mathcal{L}$ and update all trainable parameters until the loss converges or $iters$ is reached.

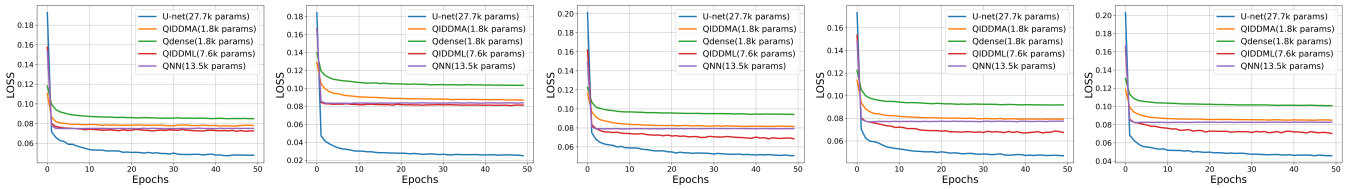
After training, the IQINN parameters θ define a consistency function $\hat{f}_\theta$ that maps any Gaussian noise $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$ directly to a denoised image $\hat{x}_0$ in one step. In practice, following Reference [20], we apply a fixed number p of denoising steps to ensure the generated images meet the desired quality threshold δ under metric $\mathbf{d}$ (e.g., FID). The full generation procedure is given in Algorithm 2.

To characterize resource usage, we derive the following complexity result:

Theorem 3.3. *[Runtime complexity of QIDDM]For an IQINNs model with K stacked entanglement blocks, N data re-uploading*

Table 2: Comparison of network costs.

Models	Qubits/Bits	Depth	Parameters	Model type	Run time Complexity	Sampling Complexity
U-net	d^2	$O(1)$ $cdepth$	$O(d^2 \cdot cdepth \cdot 2^{cdepth})$	Classical	$O(d^2 \cdot 2^{cdepth})$	$O(1) \sim O(T)$
QNN	n	$qdepth$	$2(d^2 \times n + n) + qdepth \cdot n \approx O(d^2)$	Hybrid	$O(n)$	$O(T)$
Qdense	$n \geq \log d^2$	$qdepth$	$qdepth \cdot n \approx O(n)$	Fully quantum	$O(n)$	$O(1)$
QIDDMA	n	$N \cdot L \cdot (2K+1)$	$N \cdot L \cdot (2K+1) \cdot n \approx O(n)$	Fully quantum	$O(n)$	$O(1)$
QIDDML	n	$N \cdot L \cdot (2K+1)$	$(d^2 \times n + n) + N \cdot L \cdot (2K+1) \cdot n \approx O(d^2)$	Hybrid	$O(n)$	$O(1)$



(a) Loss on label 0 (b) Loss on label 1 (c) Loss on label 2 (d) Loss on label 3 (e) Loss on label 4

Figure 3: The training loss over epochs for five classical and quantum diffusion models on the MNIST dataset, specifically for labels 0-4, is shown.

modules of length L, the run time complexity of running QIDDM on a n-qubit quantum computer is $O(L \cdot N \cdot (2K+1) \cdot n) \approx O(n)$.

Proof. As shown in Figure 1 (b), both two models of IQINNs are composed of single qubit gate R_z, Rot and two qubits gate U_{CNOT}, which means all the quantum gates used in QIDDM are NISQ device-friendly and each cost O(1) [29]. Therefore, for each layer of N data re-uploading modules, QIDDM costs run time complexity of $O(L \cdot n)$ for L angle encoding layers $S(h)$ and $O(L \cdot 2K \cdot n)$ for L trainable parameterized layers.

Parameter cost: We can further deduce that the parameters of the quantum neural network in IQINNs scale as $O(n)$. Specifically, the total number of parameters for QIMML includes additional d^2 parameters from the linear layer used for final dimensionality expansion, which is $(N+1)d^2$ fewer parameters compared to QINNs. For a trivial angle-encoded quantum neural network (QNN), twice the number of parameters is required to achieve image generation quality comparable to QIMML. On the other hand, QIMMA, which lacks the auxiliary fitting capability provided by the spectral broadening of linear layers, has a parameter count similar to that of the trivial amplitude-encoded quantum neural network (QDense). However, the image generation quality of QIMMA surpasses that of QDense. A detailed comparison of these models as well as classical counterpart (Unet) is provided in the Table 2.

4 EXPERIMENT

Experiment and Baseline Settings: We evaluate the effectiveness and efficiency of our proposed models on three benchmark datasets: MNIST, Fashion-MNIST, and EMNIST. The QIDDML model is primarily compared with a quantum neural network (QNN) based on angle encoding. To ensure a fair comparison, both QIDDML and QNN adopt nearly identical quantum architectures; however, QNN includes two additional linear layers, resulting in 13.5k parameters—nearly twice that of QIDDML's 7.6k. We also compare QIDDMA, a fully quantum neural network, with the state-of-the-art Qdense model [20], both of which feature comparable quantum structures and 1.8k parameters.

Additionally, we benchmark against a classical U-Net model [35] with a depth of 3 and an initial channel count of 8 (27.7k parameters), and an enhanced variant with 14 channels (84.1k parameters) for experiments on Fashion-MNIST. U-Net serves as a classical reference to assess the relative performance of quantum models.

For experimental configurations not specified in the original literature—such as EMNIST training for Qdense—we employ the RayTune library [22] to optimize hyperparameters across all models. Given the sensitivity of U-Net to label distributions, we adopt label-specific learning rates to enhance image generation quality and ensure a fair comparison with quantum approaches. We further compare our models with the latest quantum generative adversarial network, PQWGAN [43] (533.5k parameters), and its classical counterpart WGAN-GP [9] (899.6k parameters), using the settings reported in their respective papers.

All models are trained on a server equipped with an Intel Xeon Silver 4210R CPU (20 cores) and four NVIDIA Tesla T4 GPUs, each with 16 GB of memory.

Evaluation Metrics: We use several standard metrics to evaluate the similarity between generated and real images in the test set: Structural Similarity Index (SSIM) [45], Peak Signal-to-Noise Ratio (PSNR) [20], Fréchet Inception Distance (FID) [16], and Cosine Similarity [52]. SSIM measures the perceptual quality of generated images by evaluating structural similarity, ranging from -1 to 1, with 1 indicating perfect structural alignment. PSNR quantifies reconstruction quality by measuring the ratio of signal power to noise; higher values imply better fidelity. FID assesses the distance between feature distributions of real and generated samples, with lower scores indicating better alignment. Cosine Similarity evaluates directional similarity between feature vectors, ranging from -1 to 1, where 1 denotes perfect alignment. All scores are averaged over the test set.

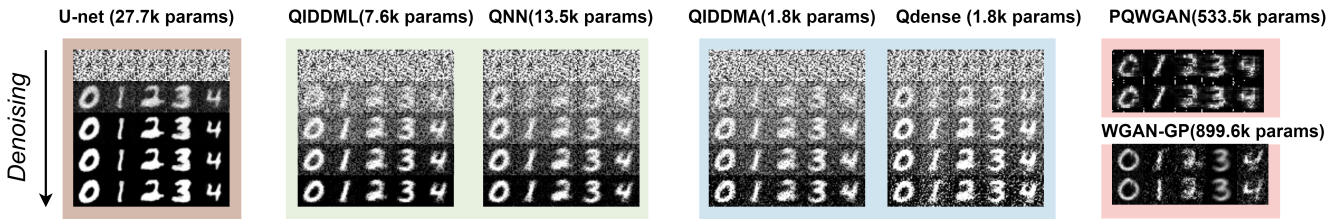


Figure 4: A comparison of the classical diffusion model (U-net), our two proposed models (QIDDML and QIAAMA), QNN, Qdense, PQWGAN, and WGAN-GP for generating 28×28 images on the MNIST dataset is presented. Only the first five labels are shown.

Table 3: Average Fid Metrics Comparison of QNN (13.5k parameters), Qdense (1.8k parameters), U-net (27.7k parameters), QGANs (533.5k parameters), GANs(899.6k parameters), and Our Proposal QIDDM (7.5k parameters) on MNIST Dataset.

Labels	0	1	2	3	4	5	6	7	8	9
PQWGAN	367.43	253.93	337.65	326.01	298.13	354.03	327.86	289.26	348.83	304.76
WGAN-GP	305.93	168.97	296.29	221.58	187.95	227.70	264.97	203.33	283.04	273.28
QNN	156.23	**71.09**	**129.19**	117.19	**97.62**	**111.78**	**120.06**	100.23	123.02	102.40
Qdense	225.50	227.11	222.28	212.28	234.33	238.81	216.50	224.89	220.35	217.55
QIDDMA	182.80	101.24	155.59	139.67	126.77	145.54	146.57	128.23	149.83	130.91
QIDDML	**154.72**	71.47	135.98	**114.73**	98.41	113.36	126.40	**93.84**	**118.53**	**99.88**
U-net(Reference)	203.69	64.00	174.74	202.50	104.89	229.70	148.70	110.39	205.91	169.36

Table 4: Average SSIM Metrics Comparison of QNN (13.5k parameters), Qdense (1.8k parameters), U-net (27.7k parameters), QGANs (533.5k parameters), GANs(899.6k parameters), and Our Proposal QIDDM (7.5k parameters) on MNIST Dataset.

Labels	0	1	2	3	4	5	6	7	8	9
PQWGAN	0.31	0.36	0.19	0.29	0.21	0.17	0.26	0.26	0.18	0.24
WGAN-GP	0.11	0.47	0.16	0.17	0.24	0.21	0.21	0.20	0.29	0.29
Qdense	0.34	0.16	0.24	0.26	0.17	0.16	0.23	0.21	0.24	0.20
QNN	**0.45**	0.30	**0.33**	0.38	0.28	0.27	0.33	0.32	0.34	0.30
QIDDMA	0.39	0.22	0.28	0.33	0.23	0.23	0.30	0.29	0.29	0.25
QIDDML	0.42	**0.49**	0.30	**0.43**	**0.35**	**0.28**	**0.36**	**0.46**	**0.38**	**0.33**
U-net(Reference)	0.46	0.62	0.37	0.36	0.37	0.27	0.45	0.51	0.36	0.43

Table 5: Average PSNR Metrics Comparison of QNN (13.5k parameters), Qdense (1.8k parameters), U-net (27.7k parameters), QGANs (533.5k parameters), GANs(899.6k parameters), and Our Proposal QIDDM (7.5k parameters) on MNIST Dataset.

Labels	0	1	2	3	4	5	6	7	8	9
PQWGAN	10.29	13.25	10.26	10.73	11.05	10.11	10.97	10.78	10.54	11.34
WGAN-GP	8.18	11.34	8.79	9.00	9.96	10.03	8.99	8.69	10.12	10.12
QNN	**10.39**	14.36	**10.56**	11.34	11.66	10.76	11.30	11.95	11.07	11.82
Qdense	7.98	7.43	7.69	7.92	7.24	7.14	7.99	7.62	7.94	7.74
QIDDMA	9.20	11.71	9.35	10.18	10.25	9.53	10.30	10.75	9.87	10.40
QIDDML	10.05	**14.76**	9.82	**11.63**	**11.75**	**10.73**	**11.33**	**12.51**	**11.30**	**12.05**
U-net(reference)	9.40	14.34	9.54	9.11	11.28	8.06	10.72	12.02	9.11	9.99

Table 6: Average Cosine similarity Metrics Comparison of QNN (13.5k parameters), Qdense (1.8k parameters), U-net (27.7k parameters), QGANs (533.5k parameters), GANs(899.6k parameters), and Our Proposal QIDDM (7.5k parameters) on MNIST Dataset.

Labels	0	1	2	3	4	5	6	7	8	9
PQWGAN	0.82	0.82	0.76	0.87	0.78	0.76	0.81	0.78	0.79	0.76
WGAN-GP	0.82	0.82	0.76	0.87	0.77	0.74	0.80	0.81	0.81	0.78
QNN	**0.88**	0.87	**0.85**	**0.87**	**0.84**	**0.81**	**0.87**	0.85	**0.87**	0.85
Qdense	0.84	0.75	0.81	0.81	0.77	0.77	0.82	0.79	0.83	0.79
QIDDMA	0.86	0.83	0.83	0.85	0.82	0.80	0.86	0.84	0.86	0.83
QIDDML	0.87	**0.88**	0.84	**0.87**	**0.84**	**0.81**	**0.87**	**0.86**	**0.87**	**0.86**
U-net(reference)	0.87	0.84	0.84	0.85	0.81	0.81	0.86	0.85	0.87	0.84

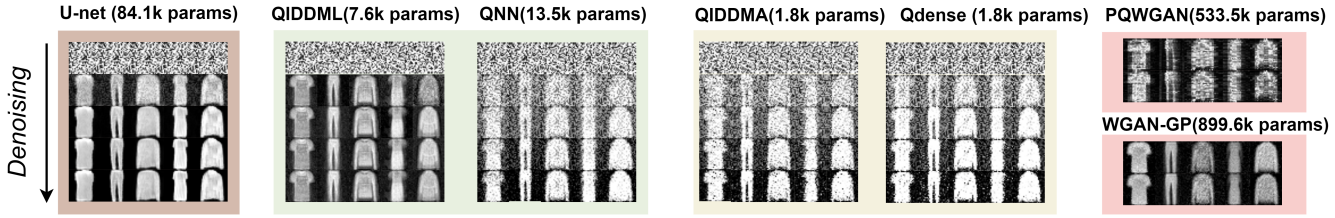


Figure 5: A comparison of the classical diffusion model (U-net), our two proposed models (QIDDML and QIAAMA), QNN, Qdense, PQWGAN, and WGAN-GP for generating 28×28 images on the fashion-MNIST dataset is presented. Only the first five labels are shown.

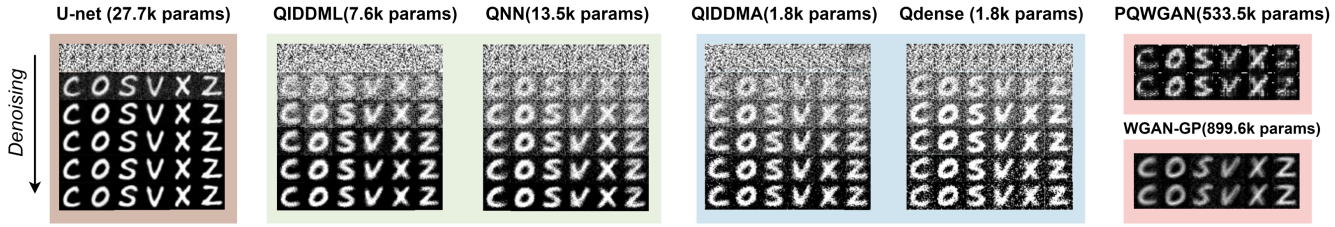


Figure 6: A comparison of the classical diffusion model (U-net), our two proposed models (QIDDML and QIAAMA), QNN, Qdense, PQWGAN, and WGAN-GP for generating 28×28 images on the E-MNIST dataset is presented. Only the first five labels are shown.

Table 7: Average Fid Metrics Comparison of QNN (13.5k parameters), Qdense (1.8k parameters), U-net (84.1k parameters), QGANs (533.5k parameters), GANs(899.6k parameters), and Our Proposal QIDDM (7.5k parameters) on fashion-MNIST Dataset.

Labels	T-shirt	Trouser	Pullover	Dress	Coat	Sandal	Shirt	Sneaker	Bag	Boot
PQWGAN	183.22	130.94	206.42	157.73	226.74	164.90	178.73	92.30	196.88	179.90
WGAN-GP	162.94	141.08	180.89	190.86	270.22	205.15	160.15	101.19	183.19	166.39
QNN	192.57	156.21	209.85	191.99	222.79	197.49	180.19	118.22	199.38	190.36
Qdense	239.87	193.52	253.96	213.67	231.00	256.88	266.56	210.66	248.53	220.64
QIDDMA	221.85	185.49	234.14	189.26	223.09	159.02	249.31	166.95	241.76	210.64
QIDDML	**92.83**	**88.74**	**105.90**	**114.41**	**102.03**	**109.45**	**129.38**	**85.12**	**132.21**	**106.45**
U-net(Reference)	180.02	123.95	151.01	194.17	187.71	89.93	182.34	111.30	180.33	191.14

Ablation Studies: To better understand the contributions of different components, we conduct several ablation comparisons: (1) QNN without the data re-uploading module is used to assess the impact of data re-uploading; (2) QNN and Qdense are compared to examine the effects of angle versus amplitude encoding; (3) classical U-Net is included as a baseline to represent classical performance; and (4) QIDDML and QIDDMA are used to compare a hybrid implicit quantum model with a linear layer against a fully quantum variant.

4.1 Experiments on Standard Datasets

MNIST 28*28. We focus on handwritten digit recognition (digits 0-9) using a dataset of 5,000 samples (approximately 500 per class), with 20% allocated to the test set.

As shown in Figure 3, all models converge within 10 epochs. Due to its large parameter size, the classical U-net model achieves the lowest convergence loss. The QIMML model follows closely, outperforming all other quantum models. Notably, QIMMA, with only 1.8k parameters, almost matches the performance of QNN, which has 7.6k parameters. In contrast, Qdense significantly underperforms. Figure 4 compares the images generated by each model. The U-net model produces clear, low-noise images, while our QIDDML model generates higher-quality images with fewer noise artifacts than QNN. QIDDMA, though slightly coarser than QNN, still produces far fewer noise artifacts than Qdense, highlighting the superior ability of quantum implicit networks to learn data distributions.

To facilitate comparison, we report the average FID, SSIM, PSNR, and Cosine Similarity scores across all 10 MNIST classes in Tables 3, 4, 5, and 6.

As shown in Table 3, which presents FID scores (lower is better), QIDDML and QNN achieve comparable performance. This similarity arises because the minor noise artifacts in QNN outputs have limited impact on FID, which primarily reflects overall distribution alignment rather than local noise. Remarkably, QIDDMA—despite having significantly fewer parameters—achieves FID scores close to those of both QIDDML and QNN, demonstrating strong generative capability under constrained model size. In contrast, the FID scores of both Qdense and the GAN-based models are substantially higher, indicating inferior image generation quality.

Table 8: Average Fid Metrics Comparison of QNN (13.5k parameters), Qdense (1.8k parameters), U-net (27.7k parameters), QGANs (533.5k parameters), GANs(899.6k parameters), and Our Proposal QIDDM (7.5k parameters) on E-MNIST Dataset.

Labels	C	O	S	V	W	X	Z
PQWGAN	394.06	377.44	346.65	365.00	325.89	345.76	358.42
WGAN-GP	232.76	299.41	247.83	239.96	243.52	242.42	276.71
QNN	151.26	185.30	156.60	140.88	144.28	157.93	147.86
Qdense	236.06	234.44	230.39	231.59	239.23	236.05	243.61
QIDDMA	185.61	201.93	184.62	175.49	173.92	176.01	194.04
QIDDML	**150.20**	**181.67**	**156.75**	**136.47**	**140.73**	**140.87**	**147.17**
U-net(Reference)	135.52	152.24	239.18	146.91	213.76	155.28	184.43

For SSIM, PSNR, and Cosine Similarity (higher is better), reported in Tables 4, 5, and 6, our QIDDML model consistently outperforms both QNN and the classical U-Net across most metrics, despite having only half the number of parameters as QNN. These results underscore the efficiency and representational strength of our model. Furthermore, QIDDMA significantly surpasses Qdense in all three metrics, even though the two models share identical parameter counts, further highlighting the advantage of our proposed improved implicit quantum architecture.

Fashion-MNIST 28*28. As shown in Figure 5, the QIDDML model demonstrates outstanding performance on the Fashion-MNIST dataset. Its generated images exhibit rich visual details, such as the collars on shirts and long sleeves, the layered structure of dresses, and even pocket contours on coats. These fine-grained features highlight QIDDML's strong ability to capture and represent complex visual patterns. In contrast, other quantum models tend to produce only coarse outlines and basic shapes, lacking the nuanced details found in QIDDML outputs.

Across all evaluation metrics, QIDDML consistently outperforms the other models, including the classical reference model U-Net. Notably, QIDDMA also surpasses Qdense in image generation quality, further demonstrating the superior expressiveness of implicit quantum neural networks (IQNNs) when modeling high-complexity data distributions. Due to space limitations, we report only the FID scores for each model in Table 7.

E-MNIST 28*28. As illustrated in Figure 6, the QIDDML model achieves the highest visual quality among all quantum models on the E-MNIST dataset, generating images with the fewest noise artifacts. Its outputs are consistently cleaner and better structured, leading to slightly superior performance compared to QNN in most cases. Furthermore, QIDDMA significantly outperforms Qdense despite both models having identical, low parameter counts. This result reinforces the effectiveness of our implicit quantum architecture under stringent resource constraints. The corresponding FID scores for each model across the selected seven E-MNIST labels are summarized in Table 8.

4.2 Test on complex data set

Beyond standard benchmarks, we evaluate QIDDM on more complex datasets: CelebA [25], MedMNIST [48], Logo_2k [44], and Fruit_360 [27]. Each label is augmented to include 100 training samples, with up to 20 labels per dataset. As shown in Figure 7, QIDDM effectively captures key image structures despite data complexity. Although the outputs are not perfectly sharp, the average SSIM between generated and real images exceeds 89% across all labels and datasets. More visual results and metric details are available on our GitHub.

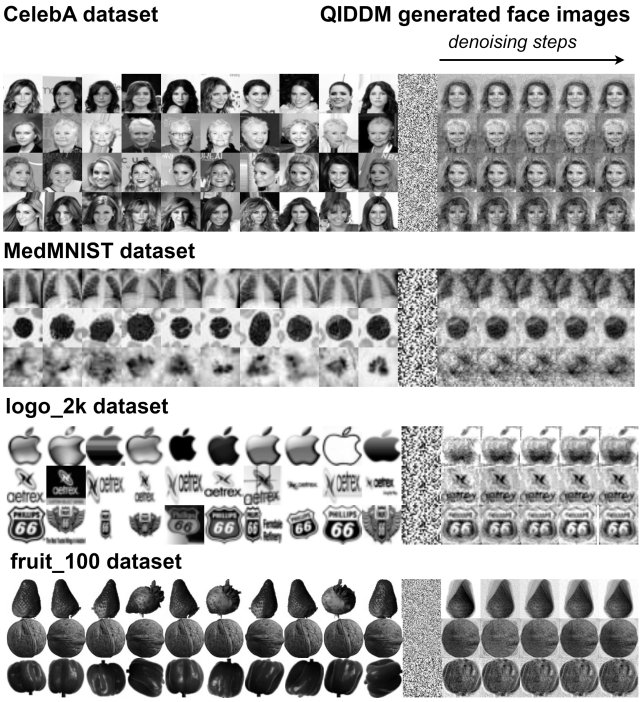


Figure 7: Complex image generation results on the CelebA, MedMNIST, logo_2k and fruit_360 dataset. Only selected attribute labels and their corresponding training samples (10 per class) are shown. Additional results can be found in the accompanying GitHub repository.

To our knowledge, this is the first application of quantum denoising diffusion models to complex image generation. Other models, such as QDense, fail to learn facial features, consistent with previous findings on its limited generalization from heterogeneous datasets. These results highlight the superior robustness and expressiveness of QIDDM.

4.3 Quantum noise analysis and test on real quantum hardware

Quantum Noise Analysis: As illustrated in Figure 8, we conduct a comprehensive robustness evaluation of all quantum models across multiple datasets and label categories under various types of

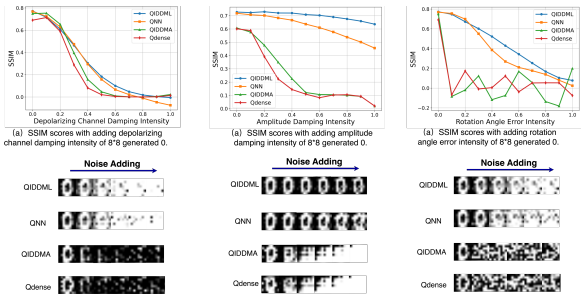


Figure 8: Comparison of SSIM values for different quantum models as the strength of depolarizing channel damping and rotation angle errors increases.

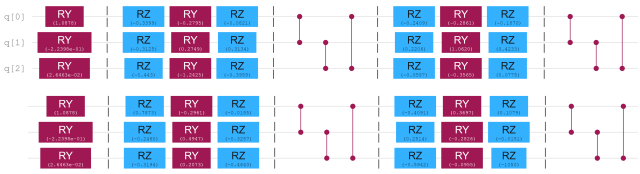


Figure 9: Quantum circuit construction of 3-qubit QIMML for generating label 0 of the 8×8 MNIST dataset

quantum noise. To ensure a fair comparison, we maintain identical quantum network architectures for QNN, QIDDML, QIDDMA, and QDense throughout all experiments. The generation of the digit "0" at an 8×8 resolution is used as a representative case study.

Since phase shift noise does not affect measurements in the computational (Z) basis or the resulting output distributions, we focus on three representative and practically impactful noise types: amplitude damping, depolarizing channel noise, and angle error noise [29], all applied uniformly across qubits. To quantify robustness, we compute SSIM scores on a shared test set while gradually increasing noise intensity in increments of 0.05 up to 1.0.

Among the three noise types, depolarizing noise proves most detrimental. All models exhibit a similar downward trend, but QIDDML and QNN maintain slightly higher SSIM scores, likely due to their classical linear layers that help absorb noise. Notably, QIDDMA surpasses QDense, suggesting that the implicit structure of IQNNs confers inherent robustness.

Under amplitude damping noise, QIDDML again demonstrates the strongest resistance. This can be attributed to its use of quantum expectation values and the stabilizing effect of data re-uploading. Although QIDDMA also outperforms QDense, both models—being heavily reliant on amplitude information—are more vulnerable to this type of noise, resulting in steeper performance degradation.

In the case of angle error noise—modeled as random perturbations to all rotation angles—QIDDMA and QDense experience rapid breakdown. Their SSIM scores collapse to near zero after just two increments and remain unstable, reflecting an inability to tolerate rotational inaccuracies. In contrast, QIDDML maintains notably higher robustness, with a significantly slower decline than QNN.

In summary, while depolarizing noise is the most damaging overall, QIDDML consistently exhibits the highest resilience across all

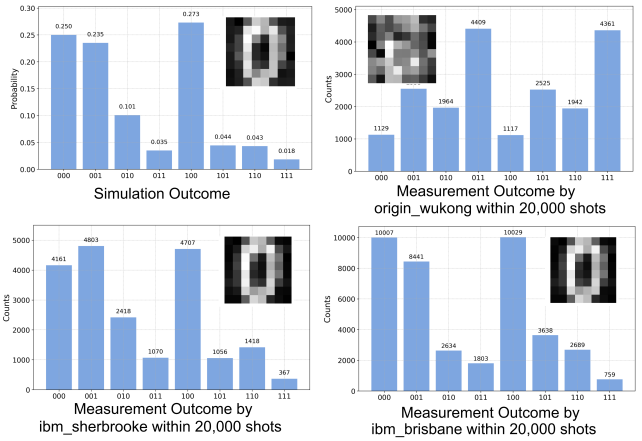


Figure 10: Generation task of 8×8 MNIST label 0 images on the simulator and the experimental results from three superconducting quantum computers: Origin_wukong, IBM_sherbrooke, and IBM_brisbane.

tested scenarios. Despite sharing sensitivity to amplitude damping with QDense, QIDDMA benefits significantly from the data re-uploading mechanism, validating the stabilizing role of our proposed IQINN architecture.

Test on real quantum hardware: We reproduce a digit "0" using QIDDM on three superconducting quantum computers. The experiment uses a 2-layer data re-uploading QIDDML model with 3 quantum bits. The quantum circuit for the theoretically trained IQINNs is shown in Figure 9, with rotation angles representing the weights. The three quantum computers tested are: Origin WuKong (72 qubits, T1 = 18.72 μ s, T2 = 1.46 μs, 20,000 shots), IBM Sherbrooke (127 qubits, T1 = 260 μs, T2 = 174.29 μs, 20,000 shots), and IBM Brisbane (127 qubits, T1 = 218.62 μs, T2 = 125.81 μs, 40,000 shots).

Figure 10 compares the theoretical generated image, the quantum circuit's final state probability distribution from the simulator, and the results from the three quantum computers. Quantum fidelity for Origin Wukong is 0.958, 0.96, and 0.99, but the lower T1 and T2 values cause noticeable deviations from the theoretical distribution, resulting in a poor resemblance to label "0". In contrast, IBM's quantum computers achieve fidelities above 0.99, with results closely matching the theoretical distribution, even with 20,000 shots. The generated images are nearly identical to the theoretical ones.

Theoretical Pauli Z values are (0.3336, 0.60446, 0.2436). Measured values are: Wukong: (0.0050, -0.2687, -0.385); Sherbrooke: (0.2703, 0.4727, 0.2452); Brisbane: (0.26795, 0.6057, 0.1442). As seen, Wukong's measurements deviate significantly from the theoretical values, causing a failure in image sampling.

5 CONCLUSION

In this paper, we propose the Quantum Implicit Denoising Diffusion Model (QIDDM), based on Quantum Implicit Neural Networks (QINNs) and the consistency model. By improving QINNs, we design two models: the high-quality image generation model, QIDDML, and the fully quantum, low-parameter model, QIMMA. Compared to recent quantum denoising diffusion models, QIDDM

demonstrates superior performance in terms of low parameter count and high image generation quality. When compared with the latest quantum generative adversarial networks and classical GANs, QIDDM achieves higher-quality image generation with less than 1% of the parameters used by these models. We further train QIDDM on the CelebA dataset for face generation, which demonstrates the potential of quantum denoising diffusion models on complex datasets. Noise anglysis and tests conducted on three different superconducting quantum computers also validate the robustness and correctness of our model. In future work, we plan to extend QIDDM to downstream tasks and explore its potential applications, such as human pose prediction.

REFERENCES

[1] J Achiam, P Liang, A Ray, and et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] MH Amin, E Andriyash, J Rolfe, and et al. 2018. Quantum Boltzmann machine. *Phys Rev X* 8, 2 (2018), 021050.

[3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[4] K Beer, D Bondarenko, T Farrelly, and et al. 2020. Training deep quantum neural networks. *Nat Commun* 11, 1 (2020), 808.

[5] M Benedetti, D Garcia-Pintos, A Perdomo-Ortiz, and et al. 2019. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Inf* 5, 1 (2019), 45.

[6] M Benedetti, E Lloyd, S Sack, and et al. 2019. Parameterized quantum circuits as machine learning models. *Quantum Sci Technol* 4, 4 (2019), 043001.

[7] J Biamonte, P Wittek, N Pancotti, and et al. 2017. Quantum machine learning. *Nature* 549, 7671 (2017), 195–202.

[8] P Braccia, Y Li, G Arce, and et al. 2021. How to enhance quantum generative adversarial learning of noisy information. *New J Phys* 23, 5 (2021), 053024.

[9] G Cai, X He, H Liu, and et al. 2019. Unsupervised domain adaptation with adversarial residual transform networks. *IEEE Trans Neural Netw Learn Syst* 31, 8 (2019), 3073–3086.

[10] Shouvanik Chakrabarti, Huang Yiming, Tongyang Li, Soheil Feizi, and Xiaodi Wu. 2019. Quantum Wasserstein generative adversarial networks. *Advances in Neural Information Processing Systems* 32 (2019).

[11] Chuangtao Chen, Qinglin Zhao, MengChu Zhou, Zhimin He, Zhili Sun, and Haozhen Situ. 2024. Quantum generative diffusion model: a fully quantum-mechanical model for generating quantum state ensemble. *arXiv preprint arXiv:2401.07039* (2024).

[12] J Chen, Y Liu, T Zhang, and et al. 2023. MOGAN: Morphologic-structure-aware generative learning from a single image. *IEEE Trans Syst Man Cybern Syst* (2023).

[13] G Cohen, S Afshar, J Tapson, and et al. 2017. EMNIST: Extending MNIST to handwritten letters. In *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*. 2921–2926.

[14] F De Falco, F Caruso, F Petruccione, and et al. 2024. Towards efficient quantum hybrid diffusion models. *arXiv preprint arXiv:2402.16147* (2024).

[15] I Gulrajani, F Ahmed, M Arjovsky, and et al. 2017. Improved training of Wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS–17)*. Curran Associates Inc., Red Hook, NY, USA, 5769–5779.

[16] M Heusel, H Ramsauer, T Unterthiner, and et al. 2017. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Adv Neural Inf Process Syst*, Vol. 30. 6626–6637.

[17] J Ho, A Jain, and P Abbeel. 2020. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 6840–6851.

[18] A Khoshaman, E Andriyash, M Benedetti, and et al. 2018. Quantum variational autoencoder. *Quantum Sci Technol* 4, 1 (2018), 014001.

[19] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[20] Michael Kölle, Gerhard Stenzel, Jonas Stein, Sebastian Zielinski, Björn Ommer, and Claudia Linnhoff-Popien. 2024. Quantum denoising diffusion models. In *2024 IEEE International Conference on Quantum Software (QSW)*. IEEE, 88–98.

[21] Y LeCun, L Bottou, Y Bengio, and et al. 1998. Gradient-based learning applied to document recognition. *Proc IEEE* 86, 11 (1998), 2278–2324.

[22] R Liaw, E Liang, R Nishihara, and et al. 2018. Tune: A research platform for distributed model selection and training. *arXiv preprint arXiv:1807.05118* (2018).

[23] H Lin, Z Liu, Q Wang, and et al. 2023. How generative adversarial networks promote the development of intelligent transportation systems: A survey. *IEEE/CAA J Autom Sinica* (2023).

[24] JG Liu and L Wang. 2018. Differentiable learning of quantum circuit Born machines. *Phys Rev A* 98, 6 (2018), 062324.

[25] Z Liu, P Luo, X Wang, and X Tang. 2015. Deep learning face attributes in the wild. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. 3730–3738.

[26] S Lloyd and C Weedbrook. 2018. Quantum generative adversarial learning. *Phys Rev Lett* 121, 4 (2018), 040502.

[27] Horea Muresan and Mihai Oltean. 2018. Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica* 10, 1 (2018), 26–42.

[28] Alexander Quinn Nichol and Prafulla Dhariwal. 2021. Improved denoising diffusion probabilistic models. In *International conference on machine learning*. PMLR, 8162–8171.

[29] Michael A Nielsen and Isaac L Chuang. 2010. *Quantum computation and quantum information*. Cambridge university press.

[30] MY Niu, B Zhang, X Wang, and et al. 2022. Entangling quantum generative adversarial networks. *Phys Rev Lett* 128, 22 (2022), 220505.

[31] M Parigi, S Martina, and F Caruso. 2023. Quantum-noise-driven generative diffusion models. *Adv Quantum Technol* 6 (2023), 2300401.

[32] J Preskill. 2018. Quantum computing in the NISQ era and beyond. *Quantum* 2 (2018), 79.

[33] A Pérez-Salinas, A Cervera-Lierta, E Gil-Fuster, and et al. 2020. Data re-uploading for a universal quantum classifier. *Quantum* 4 (2020), 226.

[34] R Rombach, A Blattmann, D Lorenz, and et al. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10684–10695.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 234–241.

[36] M Schuld and N Killoran. 2019. Quantum machine learning in feature Hilbert spaces. *Phys Rev Lett* 122, 4 (2019), 040504.

[37] M Schuld, R Sweke, and JJ Meyer. 2021. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Phys Rev A* 103, 3 (2021), 032430.

[38] J Shi, Z Luo, Y Li, and et al. 2022. Parameterized Hamiltonian learning with quantum circuit. *IEEE Trans Pattern Anal Mach Intell* 45, 5 (2022), 6086–6095.

[39] J Sohl-Dickstein, E Weiss, N Maheswaranathan, and et al. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the International Conference on Machine Learning (ICML)*. PMLR, 2256–2265.

[40] J Song, C Meng, and S Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).

[41] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. 2023. Consistency Models. In *International Conference on Machine Learning*. PMLR, 32211–32252.

[42] J Tian, Z Zhang, T Yang, and et al. 2023. Recent advances for quantum neural networks in generative learning. *IEEE Trans Pattern Anal Mach Intell* 45, 10 (2023), 12321–12340.

[43] SL Tsang, MT West, SM Erfani, and et al. 2023. Hybrid quantum–classical generative adversarial network for high-resolution image generation. *IEEE Trans Quantum Eng* 4 (2023), 1–19.

[44] Jing Wang, Weiqing Min, Sujuan Hou, Shengnan Ma, Yuanjie Zheng, Haishuai Wang, and Shuqiang Jiang. 2020. Logo-2k+: A large-scale logo dataset for scalable logo classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 6194–6201.

[45] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. 2003. Multi-scale structural similarity for image quality assessment. In *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, Vol. 2. IEEE Computer Society, 1398–1402.

[46] H Xiao, K Rasul, and R Vollgraf. 2017. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747* (2017).

[47] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2017. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1492–1500.

[48] Jiancheng Yang, Rui Shi, and Bingbing Ni. 2021. Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 191–195.

[49] Z Yu, W Liu, C Wang, and et al. 2022. Power and limitations of single-qubit native quantum neural networks. *Adv Neural Inf Process Syst* 35 (2022), 27810–27823.

[50] B Zhang, H Chen, Q Zhao, and et al. 2024. Generative quantum machine learning via denoising diffusion probabilistic models. *Phys Rev Lett* 132, 10 (2024), 100602.

[51] Jiaming Zhao, Wenbo Qiao, Peng Zhang, and Hui Gao. 2024. Quantum implicit neural representations. In *Proceedings of the 41st International Conference on Machine Learning*. 60940–60956.

[52] X Zhu, S Su, M Fu, and et al. 2018. A cosine similarity algorithm method for fast and accurate monitoring of dynamic droplet generation processes. *Sci Rep* 8 (2018), 9967.