

CoQui: A Coordinate-Conditioned Quantum Implicit Generative Adversarial Network for End-to-End Image Generation

Xue Yang^{1,2,3,4}, Rigui Zhou^{1,2*}, ShiZheng Jia^{1,2}, Dax Enshan Koh^{3,5*}, Siong Thye Goh^{4,6*},
Young-Wook Cho³, YaoChong Li^{1,2}, Xuezhi Ma³, Hongyu Chen⁷, Xin Wang⁸

¹School of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

²Research Center of Intelligent Information Processing and Quantum Intelligent Computing, Shanghai 201306, China

³Quantum Innovation Centre (Q.InC), Agency for Science, Technology and Research (A*STAR), 2 Fusionopolis Way, Innovis, #08-03, Singapore 138634, Republic of Singapore

⁴Institute of Advanced Intelligence and Computing (IAIC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore

⁵Engineering Cluster, Singapore Institute of Technology, 1 Punggol Coast Road, Singapore 828608, Republic of Singapore

⁶Lee Kong Chian School of Business, Singapore Management University (SMU), Singapore 178899, Republic of Singapore

⁷School of Computer Science and Technology, Tongji University, Shanghai 201804, China

⁸Department of Automation, Tsinghua University, Beijing 100084, P. R. China
rgzhou@shmtu.edu.cn, dax.koh@singaporetech.edu.sg, goh_siong_thye@a-star.edu.sg

Abstract

Quantum generative adversarial networks (QGANs) have recently attracted increasing attention for image generation, with the goal of using parameterized quantum circuits to model and generate image distributions. Existing approaches commonly follow a generation paradigm in which quantum-state amplitudes are mapped to pixel intensities. However, this paradigm faces two key limitations. First, existing methods typically encode pixel locations using computational-basis indices or address qubits, causing the required quantum-state dimension or number of address qubits to grow with image resolution. Second, existing methods jointly decode a large number of pixels from one or a few normalized quantum states. As a result, pixels compete for probability mass, making it difficult to control individual pixel values precisely. To address these limitations, we reformulate quantum image generation as coordinate-conditioned implicit function learning. The proposed method takes spatial coordinates and latent variables as inputs, uses a classical embedding network to generate quantum circuit parameters, and evaluates a variational quantum circuit at each coordinate location. Pixel intensities are directly read out from the expectation value of a dedicated color qubit, and a complete image is obtained by querying all spatial coordinates. This design decouples image resolution from the number of address qubits. Furthermore, our method reads out individual pixel values from separate coordinate-conditioned quantum-state evaluations, so different pixels are not required to satisfy a shared probability-normalization constraint, enabling more flexible pixel-wise modeling. To better support this coordinate-conditioned generation paradigm, we further design a specialized variational quantum circuit that introduces effective structural inductive bias for image generation. Simulated experiments on two benchmark datasets show that our method achieves better visual quality and quantitative performance than FRQI-based generation and PQWGAN baselines while using fewer qubits. Moreover, compared with the

corresponding classical baseline, the proposed model demonstrates superior generation quality and qubit efficiency.

Introduction

Recent studies have shown that QGANs have attracted increasing attention and achieved promising results in image generation tasks. Existing quantum generative adversarial network (QGAN) methods for image generation commonly adopt a generation paradigm in which pixel intensities are mapped from quantum-state amplitudes.

As one of the earliest explorations in QGAN-based image generation, Huang et al. utilized quantum-state amplitudes to represent pixel values and constructed low-resolution images through a patch-based generation strategy (Huang et al. 2021). PQWGAN follows this strategy and further extends it to the high-resolution image generation setting (Tsang et al. 2023). Although patch-based generation provides a practical way to mitigate the scalability challenge, it generates an image through multiple independently produced patches rather than modeling the entire image within a single quantum circuit. Jäger et al. recently proposed a quantum Wasserstein GAN with spatial structural inductive bias for end-to-end high-resolution image generation (Jäger, Kiwit, and Riofrío 2026). By encoding pixel positions with address qubits and pixel values with color qubits, the model generates the entire image using a single quantum circuit.

The generation paradigm based on mapping quantum-state amplitudes to pixel intensities has two major limitations. *First*, under this paradigm, the number of required qubits increases together with image resolution. *Second*, existing methods jointly recover multiple pixel values from one or a few normalized quantum states. As a result, pixels compete for probability mass, making it difficult to control individual pixel values precisely. In PQWGAN, this issue manifests as generated images whose overall brightness is lower than

*Corresponding author.

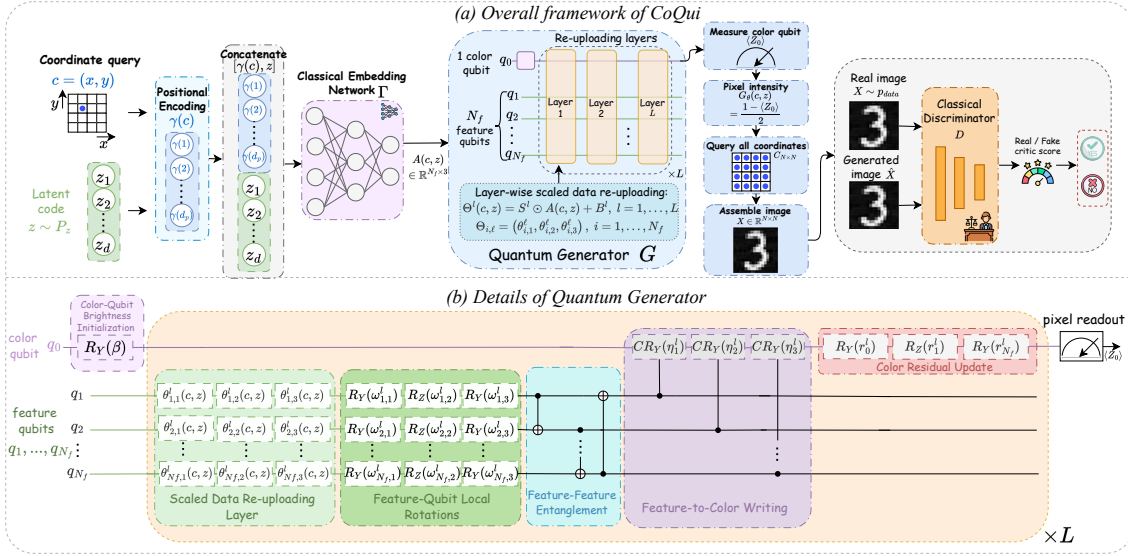


Figure 1: **Overview of the CoQui framework.** (a) **Overall architecture:** Given a spatial coordinate $c = (x, y)$ and a latent code z , CoQui utilizes a classical embedding network Γ to generate coordinate- and latent-conditioned parameters $A(c, z) \in \mathbb{R}^{N_f \times 3}$. These are transformed into layer-wise scaled parameters $\Theta_{i,\ell}$ for the quantum generator G . The pixel intensity is computed as $G(c, z) = \frac{1 - \langle Z_0 \rangle}{2}$ by measuring the Pauli-Z expectation of the color qubit. An entire image $\hat{X}$ is synthesized by querying all coordinates. (b) **Detailed quantum generator structure:** The circuit comprises one color qubit q_0 and N_f feature qubits. It first applies color-qubit brightness initialization to q_0 . Each of the L repeating layers consists of: (i) scaled data re-uploading, (ii) local feature-qubit rotations, (iii) feature-feature entanglement using a ring topology, (iv) feature-to-color writing via controlled- R_Y gates, and (v) residual color-qubit updates. Finally, q_0 gives the pixel readout.

that of the real data distribution. It can also help explain the uneven address-amplitude distribution observed in the model of Jäger et al.

In addition, many QGAN methods that we will discuss in *Related Work* use the quantum circuit only to model a classical compressed representation of the image, with the quantum circuit not directly responsible for pixel-level generation. In contrast, this work focuses on a paradigm in which the quantum circuit directly undertakes pixel-level image generation.

To address the above issues, we propose CoQui, a **coordinate-conditioned quantum implicit GAN**. The proposed framework leverages implicit neural representations (INRs) to learn a continuous mapping from spatial coordinates to signal values, thereby decoupling the number of qubits from image resolution. Moreover, unlike existing methods that jointly decode a large number of pixels from the amplitudes of one or a few quantum states, our method separately evaluates a coordinate-conditioned quantum state at each spatial coordinate and reads out the corresponding pixel value, thereby avoiding competition for probability mass among pixels. Specifically, the model takes spatial coordinates and latent variables as inputs, first maps them into the encoding parameters of a parameterized quantum circuit (PQC) through a classical network, and then directly represents continuous pixel values using quantum measurement expectations. Furthermore, this paper designs a quantum circuit architecture with structural inductive bias, where feature qubits controllably modulate a color qubit, allowing the color output to explicitly depend on the feature representation and

thereby improving the model’s capability in conditional pixel generation.

Our contributions are summarized as follows:

1. We propose CoQui (**C**oordinate-conditioned **Q**uantum **I**mplicit GAN), which formulates image generation as a continuous mapping from spatial coordinates to pixel intensities, rather than relying on the conventional amplitude-mapping paradigm. This design decouples the number of qubits from image resolution and mitigates the inter-pixel competition for probability mass inherent in conventional amplitude-based representations.
2. We design a quantum circuit with structural inductive bias, in which feature qubits controllably modulate a color qubit. This mechanism makes the color generation process explicitly dependent on the input feature representation, thereby enhancing the modeling capability of the quantum circuit.
3. We conduct systematic simulation experiments on multiple image generation benchmark datasets. Experimental results show that, compared with existing QGAN methods based on the amplitude-mapping paradigm, CoQui achieves better qualitative and quantitative performance and is able to generate images effectively with fewer qubits.

Related Work

Quantum Implicit Neural Representations and Image Generation

In recent years, implicit neural representations (INRs) have attracted substantial attention because of their ability to represent high-quality signals by learning mappings from continuous coordinates to signal values (Mescheder et al. 2019; Sitzmann et al. 2020; Park et al. 2019). Unlike conventional discrete-grid representations (Gonzalez 2009), INRs model images (Chen, Liu, and Wang 2021; Cao et al. 2023), audio (Sitzmann et al. 2020; Su, Chen, and Shlizerman 2022), and three-dimensional scenes (Zhao et al. 2024; Müller et al. 2022) as continuous functions, from which signal values can be recovered through coordinate queries. Inspired by this idea, researchers have begun to investigate Quantum Implicit Neural Representations (QINRs) (Zhao et al. 2024) built with parameterized quantum circuits (PQCs), aiming to exploit the potential of quantum circuits for high-dimensional nonlinear function approximation and Fourier-feature representation (Pérez-Salinas et al. 2020; Schuld, Sweke, and Meyer 2021).

Existing studies on QINRs mainly use them as continuous representers for conditional signals, with applications in deterministic mapping tasks such as image reconstruction (Zhao et al. 2024; Eren 2026; Wang, Theobalt, and Golyanik 2026), compression (Fujihashi and Koike-Akino 2026), and super-resolution (Jin, Singh, and Merz Jr 2025; Zhao et al. 2024). Although Zhang et al. proposed OQIDDM (Zhang et al. 2025), a QINR-based quantum diffusion model that introduces QINR into generative modeling, their method relies on amplitude encoding, which causes the required number of qubits to increase with image resolution. Thus, while QINRs show considerable promise for coordinate-dependent function representation, their integration with adversarial distribution learning remains insufficiently explored. To address this gap, this work investigates the feasibility of using QINR for end-to-end image distribution modeling within a Generative Adversarial Network (GAN) framework (Goodfellow et al. 2020).

Quantum Generative Adversarial Networks for Image Generation

GANs learn real data distributions through a minimax game between a generator and a discriminator. To alleviate training instability and mode collapse, WGAN (Arjovsky, Chintala, and Bottou 2017) employs the Wasserstein distance to provide a smoother optimization objective, while WGAN-GP (Gulrajani et al. 2017) further introduces a gradient penalty to enforce the Lipschitz constraint and improve training stability. This objective has also been adopted by many QGAN models to stabilize adversarial training. Quantum Generative Adversarial Networks (QGANs) extend adversarial generative modeling to the quantum computing framework. Early studies established the theoretical foundations of quantum adversarial learning (Lloyd and Weedbrook 2018) and further used parameterized quantum circuits to implement trainable quantum generative models (Dallaire-Demers and Killoren 2018), with applications to numerical distribution modeling and quantum-state preparation (Zoufal, Lucchi, and

Woerner 2019).

More recently, QGANs have been applied to image generation. However, high-dimensional visual data impose substantial requirements on both the number of qubits and circuit depth. Existing methods therefore typically rely on dimensionality reduction or patch-based generation to alleviate scalability issues. Dimensionality-reduction approaches first compress images into a low-dimensional latent space using PCA or an encoder, then use a quantum generator to model the low-dimensional representation, and finally recover image dimensions through classical post-processing (Chu et al. 2023; Silver et al. 2023; Veličković et al. 2024; Chang et al. 2024; Thomas, Youel, and Jose 2025). Patch-based approaches divide an image into multiple local patches and use one or more quantum generators to synthesize the image blocks, thereby avoiding direct modeling of the full high-dimensional image (Huang et al. 2021; Tsang et al. 2023; Yang et al. 2026b). These methods reduce the training difficulty of quantum models, but they also weaken the direct role of the quantum generator in the complete image generation process. In particular, when classical decoders or multiple local generators are used, part of the generation capability may come from classical modules or task-specific structures.

To further improve the scalability of quantum image generation, recent studies have begun to explore end-to-end full-image quantum generators. Jäger et al. proposed a quantum Wasserstein GAN with a structural inductive bias (Jäger, Kiwit, and Riofrío 2026). Their design follows the idea of an FRQI-like address-color image representation and generates full-resolution images without relying on dimensionality reduction or patch-based generation. Meanwhile, Yang et al. improved the amplitude decoding and prior-mapping mechanisms of the PQWGAN model, enabling end-to-end quantum image generation (Yang et al. 2026a).

We observe that most existing QGAN methods rely on quantum-state amplitudes to represent pixel intensities or feature values. This mechanism is constrained by the global normalization of quantum states and inevitably introduces numerical coupling among pixels, which weakens the model’s ability to independently control local signal values. To address this limitation, this work explores an implicit quantum representation based on coordinate queries, where spatial coordinates and latent variables jointly drive a parameterized quantum circuit and pixel intensities are directly extracted from measurement expectations of observables. This formulation provides a more flexible route toward end-to-end image distribution modeling.

Method

Problem Formulation

This work proposes CoQui, a QGAN framework based on Quantum Implicit Neural Representation (QINR). Instead of generating an image by assigning pixel intensities to quantum-state amplitudes, CoQui reformulates image generation as a continuous implicit mapping conditioned jointly on spatial coordinates and a latent variable.

CoQui represents a grayscale image as $X \in [0, 1]^{H \times W \times 1}$ and learns a coordinate-conditioned generator $G_{\Phi}(c, z) \in$

$[0, 1]$. Here, $c = (x, y) \in \mathcal{C} \subset [0, 1]^2$ is a normalized pixel coordinate, $z \in \mathbb{R}^{d_z}$ is a latent code shared across all coordinates of one image, and Φ denotes the trainable parameters. A complete image is obtained by evaluating the generator over $\mathcal{C}_{H,W} = \{(i/W, j/H) \mid 0 \leq i < W, 0 \leq j < H\}$.

Conditional Input to the Quantum Implicit Generator

To improve the ability of coordinate inputs to represent local details and high-frequency structures, the two-dimensional coordinate $c = (x, y)$ is first transformed by positional encoding. Given the number of frequencies K , the encoding function is defined as

$$\gamma(c) = \left[x, y, \{\sin(2^k x), \cos(2^k x), \sin(2^k y), \cos(2^k y)\}_{k=0}^{K-1} \right]. \quad (1)$$

with encoded coordinate dimension $d_\gamma = 2 + 4K$. The encoded coordinate feature is concatenated with the latent variable $z \in \mathbb{R}^{d_z}$:

$$u(c, z) = [\gamma(c), z] \in \mathbb{R}^{d_\gamma + d_z}. \quad (2)$$

This vector is fed into a classical embedding network Γ composed of fully connected layers and ReLU activations. The network outputs $3N_f$ values, where N_f is the number of feature qubits. A hyperbolic tangent nonlinearity and an angle scaling factor α are then applied to obtain the base rotation angles:

$$a(c, z) = \alpha \tanh(\Gamma(u(c, z))) \in \mathbb{R}^{3N_f}. \quad (3)$$

The vector is reshaped as

$$A(c, z) = \text{Reshape}(a(c, z)) \in \mathbb{R}^{N_f \times 3}. \quad (4)$$

For the i -th feature qubit, the three base angles are

$$A_i(c, z) = (a_{i,x}(c, z), a_{i,y}(c, z), a_{i,z}(c, z)), \quad i = 1, \dots, N_f. \quad (5)$$

which are used for R_X , R_Y , and R_Z input rotations each.

Quantum Implicit Generator Circuit

The quantum generator consists of $N_q = N_f + 1$ qubits. The first qubit q_0 is designated as the color qubit for pixel readout, while the remaining qubits $q_1, \dots, q_{N_f}$ are feature qubits that encode coordinate- and latent-dependent implicit features. The circuit is initialized from the all-zero state.

Color-Qubit Brightness Initialization Initially, a trainable R_Y rotation is applied to the color qubit:

$$U_{\text{bias}} = R_Y^{(q_0)}(\beta), \quad (6)$$

where β is a trainable brightness-bias parameter. To give the model a reasonable initial pixel-brightness distribution, β is initialized according to a preset mean pixel value $\mu_0 \in [0, 1]$:

$$\beta_0 = \arccos(1 - 2\mu_0). \quad (7)$$

Ignoring subsequent feature-to-color writing operations, the Pauli- Z expectation of the color qubit is $\langle Z_0 \rangle = \cos \beta_0$, and the corresponding initial pixel value is

$$\frac{1 - \langle Z_0 \rangle}{2} = \frac{1 - \cos \beta_0}{2} = \mu_0. \quad (8)$$

This initialization explicitly controls the initial average brightness of the generator and improves early-stage training stability.

Scaled Data Re-uploading Layer The quantum circuit comprises L data re-uploading layers. Rather than injecting the same base angles $A(c, z)$ identically at every layer, CoQui equips each layer l with trainable scale and bias parameters: $S_l \in \mathbb{R}^{N_f \times 3}$, $B_l \in \mathbb{R}^{N_f \times 3}$. The angles injected at layer l are obtained through the element-wise affine transformation

$$\Theta_l(c, z) = S_l \odot A(c, z) + B_l, \quad (9)$$

where $\odot$ denotes element-wise multiplication. Specifically, for feature qubit q_i and rotation axis $r \in \{x, y, z\}$,

$$\theta_{i,r}^l(c, z) = s_{i,r}^l a_{i,r}(c, z) + b_{i,r}^l. \quad (10)$$

Here, $s_{i,r}^l$ controls the strength of the coordinate-latent signal, whereas $b_{i,r}^l$ provides an input-independent angular offset. The three angles associated with q_i are grouped as

$$\Theta_{i,l}(c, z) = (\theta_{i,1}^l(c, z), \theta_{i,2}^l(c, z), \theta_{i,3}^l(c, z)). \quad (11)$$

This encoding acts on all feature qubits:

$$U_{\text{enc}}^l = \prod_{i=1}^{N_f} R_Z^{(q_i)}(\theta_{i,z}^l(c, z)) R_Y^{(q_i)}(\theta_{i,y}^l(c, z)) R_X^{(q_i)}(\theta_{i,x}^l(c, z)). \quad (12)$$

This layer-dependent reparametrization lets each layer emphasize, attenuate, or shift components of the shared coordinate-latent representation, enhancing the PQC's expressive capacity for implicit function modeling.

Feature-Qubit Local Rotations After input encoding, each feature qubit receives trainable local rotations with parameters $W_l \in \mathbb{R}^{N_f \times 3}$ where $W_l^i = (w_{i,1}^l, w_{i,2}^l, w_{i,3}^l)$ for the i -th feature qubit. The local rotation unit is

$$U_{\text{loc}}^l = \prod_{i=1}^{N_f} R_Y^{(q_i)}(w_{i,3}^l) R_Z^{(q_i)}(w_{i,2}^l) R_Y^{(q_i)}(w_{i,1}^l). \quad (13)$$

This component provides input-independent trainable quantum transformations after coordinate-latent encoding.

Feature-Feature Entanglement To model correlations among feature qubits, CoQui introduces a ring CNOT entanglement pattern:

$$U_{\text{ent}}^l = \prod_{i=1}^{N_f} \text{CNOT}(q_i, q_{i+1}), \quad q_{N_f+1} \equiv q_1. \quad (14)$$

Feature-to-Color Writing The core circuit design writes information from feature qubits to the color qubit through controlled- R_Y gates. For layer l and feature qubit q_i , let $\eta_i^l \in \mathbb{R}$ denote the trainable color-writing parameter. The feature-to-color operation is $\text{CRY}_{q_i \rightarrow q_0}(\eta_i^l)$, where q_i is the control qubit and q_0 is the target qubit. The complete writing module at layer l is

$$U_{\text{write}}^l = \prod_{i=1}^{N_f} \text{CRY}_{q_i \rightarrow q_0}(\eta_i^l). \quad (15)$$

This module makes the color-qubit state depend explicitly on feature qubits that encode coordinate and latent information. Thus, feature qubits form implicit spatial features, while the color qubit acts as the measurable pixel-output channel.

Color Residual Update After each feature-to-color writing module, CoQui applies a trainable residual update on the color qubit:

$$U_{\text{res}}^l = R_Y^{(q_0)}(r_{l,3})R_Z^{(q_0)}(r_{l,2})R_Y^{(q_0)}(r_{l,1}), \quad (16)$$

where $r_l = (r_{l,1}, r_{l,2}, r_{l,3})$ denotes the color residual parameters at layer l . This module provides additional local degrees of freedom for the readout qubit, allowing pixel intensities to be accumulated, adjusted, and refined across multiple re-uploading layers.

Combining the above modules, the l -th quantum transformation is

$$U_l(c, z) = U_{\text{res}}^l U_{\text{write}}^l U_{\text{ent}}^l U_{\text{loc}}^l U_{\text{enc}}^l(c, z). \quad (17)$$

For coordinate $c = (x, y)$ and latent variable z , the complete PQC outputs

$$|\psi_{\Theta}(c, z)\rangle = U_L(c, z) \cdots U_2(c, z) U_1(c, z) U_{\text{bias}} |\psi_0\rangle. \quad (18)$$

The generator parameters are $\Phi = \{s^l, b^l, W^l, \eta^l, r^l\}_{l=1}^L$.

Quantum Measurement and Pixel Readout

After L layers of quantum evolution, CoQui measures only the Pauli- Z expectation of the color qubit:

$$m_{\Phi}(c, z) = \langle Z_0 \rangle = \langle \psi_{\Phi}(c, z) | Z_0 | \psi_{\Phi}(c, z) \rangle. \quad (19)$$

The expectation is then mapped linearly to a normalized grayscale pixel intensity:

$$G_{\Phi}(c, z) = \frac{1 - m_{\Phi}(c, z)}{2}. \quad (20)$$

Since $m_{\Phi}(c, z) \in [-1, 1]$, the output satisfies $G_{\Phi}(c, z) \in [0, 1]$. This design uses a physical observable of the color qubit directly as the pixel output, making the quantum circuit itself responsible for pixel-intensity modeling.

Given a latent variable z , a full image is generated by querying all coordinates in $\mathcal{C}_{H,W}$:

$$\hat{X}_{j,i} = G_{\Phi}\left(\frac{i}{W}, \frac{j}{H}, z\right), 0 \leq i < W, 0 \leq j < H. \quad (21)$$

yielding

$$\hat{X} = G_{\Phi}(z) \in [0, 1]^{H \times W \times 1}. \quad (22)$$

Training Objective

Following PQWGAN (Tsang et al. 2023), CoQui is trained using the Wasserstein GAN objective with gradient penalty. The critic distinguishes real images from generated ones, while the generator learns to produce realistic images. Further implementation details of the classical embedding network and the critic are provided in Appendix 1.

Experiments

Experimental Setup

We evaluate CoQui on MNIST and Fashion-MNIST at the standard 28×28 resolution. The training dataset consists of 1000 real images in each experiment. The model is trained with Adam for 1000 epochs with a batch size of 5. Following the setting of Jäger et al., the learning rates of the generator and discriminator are set to 1×10^{-3} and 1×10^{-4} , respectively. The quantum generator used in the experiments contains one color qubit and $N_f = 4$ feature qubits, and the circuit depth is set to $r = 20$ re-uploading layers. The hyperparameters of the other classical components are reported in Appendix 1. All experiments are performed in simulation on a machine with an AMD Ryzen 9 9950X3D 16-Core Processor at 4.30 GHz and 64 GB RAM. All models are implemented in JAX.

Baselines. We compare CoQui with two representative quantum image generation methods, PQWGAN and the end-to-end Quantum Wasserstein QGAN proposed by Jäger et al., both of which adopt the amplitude-mapping paradigm. Experiments are conducted under two evaluation settings. We first compare the methods on the standard multi-class benchmark following the protocol of Jäger et al. We then perform per-class experiments, where each method is trained independently on each class using the same training protocol and data budget. This complementary evaluation assesses both the ability to model a heterogeneous image distribution and the generation quality under matched training conditions. In addition, we implement a classical counterpart of CoQui to isolate the contribution of the proposed quantum generator. All baseline methods are reimplemented by us and optimized for efficient classical simulation. The code will be made publicly available upon acceptance.

Evaluation Metric. We evaluate generated samples using four complementary metrics: FID, classifier-based diversity metrics, feature-space precision and recall, and auxiliary low-level visual statistics. **FID** measures the overall discrepancy between real and generated distributions (Jäger, Kiwit, and Riofrío 2026). To evaluate class diversity, a pre-trained MNIST/Fashion-MNIST classifier is used to obtain the predicted class distribution $p_g(y)$, following the classifier-based evaluation idea of Inception Score and Mode Score (Salimans et al. 2016; Che et al. 2016). We report **Entropy** as $H_{\text{norm}}(p_g) = -\frac{1}{\log C} \sum_{c=1}^C p_g(c) \log p_g(c)$, where $C = 10$, and **JSD** between $p_g(y)$ and the uniform distribution $u(y)$, defined as $\text{JSD}(p_g \| u) = \frac{1}{2} \text{KL}(p_g \| m) + \frac{1}{2} \text{KL}(u \| m)$ with $m = \frac{1}{2}(p_g + u)$ (Lin 1991). We further use **P@5**

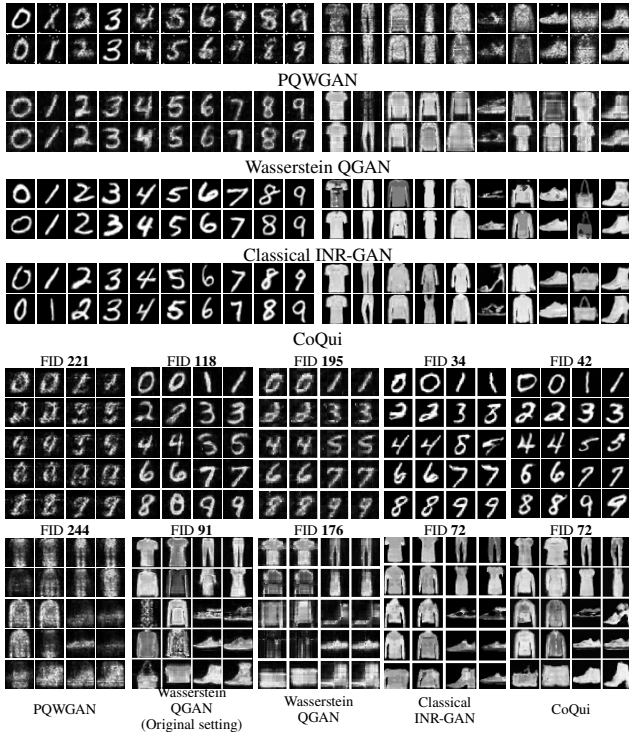


Figure 2: Qualitative comparison on MNIST and Fashion-MNIST. Top: samples from each class for each method. Bottom: representative grids over ten classes, with FID scores reported above each grid.

and **R@5** to measure sample fidelity and distribution coverage in the Inception feature space, following the improved precision and recall metric for generative models (Kynkäänniemi et al. 2019). Finally, **Brightness** and **Sobel** are defined as $D_{\text{bright}} = |\text{mean}(x_g) - \text{mean}(x_r)|$ and $D_{\text{sobel}} = |\text{mean}(\text{Sobel}(x_g)) - \text{mean}(\text{Sobel}(x_r))|$, respectively, where the Sobel operator is used to estimate image gradient magnitude (Sobel and Feldman 1968). Lower values are better for FID, JSD, Brightness, and Sobel, whereas higher values are better for Entropy, P@5, and R@5.

Dataset	Method	0	1	2	3	4	5	6	7	8	9
MNIST	PQWGAN	210	137	223	203	191	187	179	179	195	199
	Wasserstein QGAN	214	122	203	174	189	175	184	172	193	197
	Classical INR-GAN	20	16	22	14	21	19	16	20	19	19
	CoQui	17	17	26	17	19	21	17	25	19	16
	Wasserstein QGAN (Original setting)										
Fashion-MNIST	PQWGAN	241	97	204	175	213	210	192	219	275	230
	Wasserstein QGAN	167	112	164	183	162	169	165	246	184	176
	Classical INR-GAN	53	25	54	44	56	58	61	43	47	39
	CoQui	47	28	45	40	41	62	52	43	55	38
	Wasserstein QGAN (Original setting)										

Table 1: Comparison on MNIST and Fashion-MNIST. Best results are in bold.

Main Results

Figure 2 shows representative generated samples on class-wise subsets and on the full dataset containing ten classes.

Method	Qubits	Number of Quantum Circuits
PQWGAN	192	32
Wasserstein QGAN	11	1
CoQui	5	1

Table 2: Comparison of quantum resource requirements across different QGAN models.

The FID scores for the full-dataset setting are directly annotated in the figure, while the class-wise FID scores are reported in Table 1. On the class-wise subsets, CoQui generates samples with clearly recognizable category structure and good visual quality. This is especially evident on the more complex Fashion-MNIST dataset, where CoQui still produces visually meaningful samples. In contrast, the two amplitude-mapping based QGAN models generate noticeably blurrier samples. This is mainly due to the pixel coupling introduced by amplitude encoding: pixels share normalized amplitudes and are therefore difficult to adjust independently, which limits the representation of fine-grained features. Compared with the Classical INR-GAN of a similar parameter scale, CoQui achieves the best or tied-best FID scores on four MNIST classes and seven Fashion-MNIST classes. This indicates that CoQui retains strong generative capability under a low quantum-resource budget, highlighting the potential of quantum generative models for image generation. The full-dataset setting with ten classes is more challenging than class-wise generation because it involves a more complex data distribution; accordingly, all methods show degraded performance. Nevertheless, PQWGAN gives the weakest visual results, followed by Wasserstein QGAN trained under the same setting, while Wasserstein QGAN trained with the original 6K-sample setting shows some improvement. During our reproduction of Wasserstein QGAN, we observed that the original work employs a post-selection strategy for visualization: for each noise pattern, 500 candidate samples are generated, and the sample with the smallest Euclidean distance to their mean is selected for display. We do not apply this selection procedure and instead directly visualize randomly generated samples. Compared with Classical INR-GAN, CoQui achieves comparable overall visual quality and shows advantages in local detail modeling.

In addition, as shown in Table 2, CoQui uses only five qubits and one quantum circuit, giving it the lowest quantum-resource cost among all compared methods. The qubit requirement of amplitude-mapping based methods typically grows with image resolution, whereas the qubit count of CoQui is not directly tied to image resolution. CoQui therefore provides better resource efficiency and scalability. The training convergence behavior is analyzed in Appendix 2.

Ablation Study 1: circuit architecture ablation

We further conduct ablation studies to examine the contribution of the proposed circuit architecture and the color-qubit design. Figure 3 provides representative MNIST samples for these variants, while Tables 3 and 4 report the corresponding quantitative results.

Circuit design ablation Table 3 compares CoQui using the proposed feature-to-color circuit with a hardware-efficient

Metric	CoQui with hardware-efficient circuit	CoQui with our proposed circuit
FID ↓	41.41 ± 6.10	40.15 ± 3.45
JSD ↓	0.00858 ± 0.00333	0.00411 ± 0.00128
Entropy ↑	0.9856 ± 0.0055	0.9930 ± 0.0021
P@5 ↑	0.489 ± 0.081	0.505 ± 0.061
R@5 ↑	0.485 ± 0.250	0.467 ± 0.108
Brightness ↓	0.0065 ± 0.0046	0.0058 ± 0.0051
Sobel ↓	0.0200 ± 0.0000	0.0073 ± 0.0033

Table 3: Quantitative comparison between CoQui with the hardware-efficient circuit and CoQui with our proposed circuit. Best results are shown in bold.

ansatz. The proposed circuit improves most metrics, reducing FID from 41.41 to 40.15 and JSD from 0.00858 to 0.00411, while increasing entropy from 0.9856 to 0.9930 and P@5 from 0.489 to 0.505. It also yields lower brightness and Sobel discrepancies, indicating better low-level image statistics and sharper local structures. The hardware-efficient circuit gives a slightly higher R@5, but its weaker FID, JSD, and edge statistics suggest poorer overall distributional alignment. These results show that the structured feature-to-color modulation is important for stable and faithful image generation.

Core-component ablation Figure 3(b) and Table 4 evaluate the main color-qubit components. Full CoQui achieves the best JSD, entropy, brightness discrepancy, and Sobel discrepancy, showing the most balanced match to the real distribution. Removing the color-qubit residual or the color-bias initialization can slightly improve FID, but both variants worsen low-level statistics, especially edge consistency. For example, removing the residual increases the Sobel discrepancy from 0.0073 to 0.0370. Removing the scaled data re-uploading causes the largest degradation, increasing FID from 40.15 to 56.54, which indicates that layer-wise input scaling is important for effectively modulating coordinate- and noise-dependent quantum features. Overall, these ablations suggest that the complete color-qubit design is not optimized for a single metric only, but provides the best trade-off across distributional fidelity, sample diversity, brightness, and local structure.

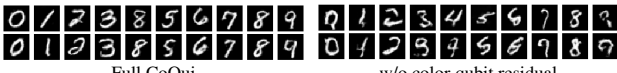


CoQui with hardware-efficient ansatz



CoQui with our proposed circuit

(a) Circuit ablation



Full CoQui

w/o color-qubit residual



w/o color-qubit bias initialization

w/o scaled data re-uploading

(b) Core-component ablation

Figure 3: Representative MNIST samples for (a) CoQui and its hardware-efficient variant, and (b) the CoQui baseline and variants without the color residual, color-bias initialization, or scaled data re-uploading.

Metric	Full CoQui	CoQui w/o color-qubit residual	CoQui w/o color-bias initialization	CoQui w/o scaled data re-uploading
FID ↓	40.15 ± 3.45	38.82 ± 1.52	36.81 ± 3.30	56.54 ± 3.63
JSD ↓	0.00411 ± 0.00128	0.00493 ± 0.00502	0.00560 ± 0.00079	0.00439 ± 0.00063
Entropy ↑	0.9930 ± 0.0021	0.9915 ± 0.0087	0.9903 ± 0.0015	0.9924 ± 0.0009
Brightness ↓	0.0058 ± 0.0051	0.0133 ± 0.0015	0.0084 ± 0.0045	0.0091 ± 0.0033
Sobel ↓	0.0073 ± 0.0033	0.0370 ± 0.0200	0.0142 ± 0.0171	0.0182 ± 0.0036

Table 4: Quantitative core-component ablation results for CoQui. Best results are shown in bold.

Ablation Study 2: Model Capacity Analysis

Figure 4 and Table 5 evaluate the effect of model capacity from two aspects: the number of feature qubits and the circuit depth. Since the computational cost of quantum circuit simulation grows rapidly with both the number of qubits and the circuit depth, we first identify a suitable number of feature qubits under a fixed circuit depth, and then compare different circuit depths based on this setting.

Ablation Study	Setting	FID ↓	Class JSD ↓	Class Entropy ↑
Feature qubits	$N_f = 2$	39.53	0.0045	0.9923
	$N_f = 3$	41.56	0.0157	0.9739
	$N_f = 4$	42.59	0.0032	0.9944
	$N_f = 5$	37.88	0.0202	0.9677
	$N_f = 6$	44.65	0.0062	0.9895
Circuit depth	10 layers	39.07	0.0032	0.9945
	20 layers	42.59	0.0032	0.9944
	30 layers	37.90	0.0079	0.9863
	40 layers	38.73	0.0082	0.9853

Table 5: Quantitative capacity comparison on MNIST. Top: varying feature qubits at $r = 20$; bottom: varying circuit depth with $N_f = 4$. Best values are bold.

Effect of Model Capacity As shown in Figure 4(a), increasing N_f from 2 to 4 clearly improves digit shape and edge quality. When $N_f = 5$, the visual quality becomes close to that of $N_f = 4$ while $N_f = 6$ shows slight degradation, e.g., parts of the strokes of digit “8” in the first row become sticky and connected, and the shape of digit “9” is also slightly distorted. Quantitatively, the fixed circuit depth of $r = 20$, $N_f = 5$ achieves the lowest FID of 37.88, but $N_f = 4$ gives the best Class JSD and Class Entropy (Table 5), so we adopt $N_f = 4$ for the subsequent depth experiments.

Figure 4(b) shows 10- and 20-layer circuits perform comparably, while 30-40 layers show mild degradation in both visual quality and class JSD/Entropy. This suggests shallow-to-moderate depth already suffices, with deeper circuits adding optimization difficulty rather than expressivity.

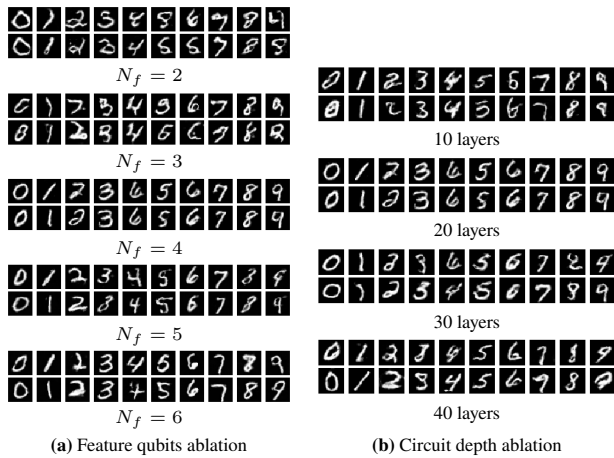


Figure 4: Representative MNIST samples for two capacity studies. (a) Different numbers of feature qubits at a fixed circuit depth of $r = 20$. (b) Different circuit depths using four feature qubits.

Conclusion and Discussion

In this paper, we propose CoQui, an end-to-end GAN framework for image generation based on quantum implicit neural representations, reformulating quantum image generation as a coordinate-conditioned implicit function learning problem. This decouples image resolution from the number of address qubits and alleviates the inter-pixel coupling induced by conventional amplitude-to-pixel mapping. We further design a quantum generator with structural inductive biases tailored to this formulation, enhancing model expressiveness. Simulation results show that CoQui achieves better visual quality and quantitative performance than amplitude-mapping-based QGAN baselines while using fewer qubits, and achieves competitive generative performance relative to a classical baseline of comparable parameter scale. The current evaluation is limited to classical simulation on standard-resolution grayscale benchmarks, without accounting for realistic hardware noise, finite-shot effects, or device connectivity. Future work will extend CoQui to higher-resolution and color image generation while improving scalability, hardware compatibility, and noise robustness toward deployment on real quantum devices.

References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein Generative Adversarial Networks. In Precup, D.; and Teh, Y. W., eds., *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, 214–223. PMLR.

Cao, J.; Wang, Q.; Xian, Y.; Li, Y.; Ni, B.; Pi, Z.; Zhang, K.; Zhang, Y.; Timofte, R.; and Van Gool, L. 2023. CioSR: Continuous Implicit Attention-in-Attention Network for Arbitrary-Scale Image Super-Resolution. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1796–1807.

Chang, S. Y.; Thanasilp, S.; Saux, B. L.; Vallecorsa, S.; and Grossi, M. 2024. Latent style-based quantum GAN for high-quality image generation. *arXiv preprint arXiv:2406.02668*.

Che, T.; Li, Y.; Jacob, A. P.; Bengio, Y.; and Li, W. 2016. Mode Regularized Generative Adversarial Networks. *arXiv preprint arXiv:1612.02136*.

Chen, Y.; Liu, S.; and Wang, X. 2021. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8628–8638.

Chu, C.; Skipper, G.; Swamy, M.; and Chen, F. 2023. IQ-GAN: Robust quantum generative adversarial network for image synthesis on NISQ devices. In *ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 1–5. IEEE.

Dallaire-Demers, P.-L.; and Killoran, N. 2018. Quantum generative adversarial networks. *Phys. Rev. A*, 98: 012324.

Eren, S. M. 2026. Implementation of Quantum Implicit Neural Representation in Deterministic and Probabilistic Autoencoders for Image Reconstruction/Generation Tasks. *arXiv preprint arXiv:2603.06755*.

Fujihashi, T.; and Koike-Akino, T. 2026. Quantum Implicit Neural Compression. In Ali, S.; Chicano, F.; and Moraglio, A., eds., *Quantum Computing and Artificial Intelligence*, 60–69. Cham: Springer Nature Switzerland. ISBN 978-3-032-15931-1.

Gonzalez, R. C. 2009. *Digital image processing*. Pearson education india.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. 2017. Improved training of Wasserstein GANs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, 5769–5779. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.

Huang, H.-L.; Du, Y.; Gong, M.; Zhao, Y.; Wu, Y.; Wang, C.; Li, S.; Liang, F.; Lin, J.; Xu, Y.; et al. 2021. Experimental quantum generative adversarial networks for image generation. *Physical Review Applied*, 16(2): 024051.

Jin, H.; Singh, G.; and Merz Jr, K. M. 2025. QFGN: A Quantum Approach to High-Fidelity Implicit Neural Representations. *arXiv preprint arXiv:2504.19053*.

Jäger, J.; Kiwit, F. J.; and Riofrío, C. A. 2026. Scaling quantum machine learning without tricks: full-resolution and diverse image generation. *Quantum Science and Technology*, 11(3): 035042.

Kynkäänniemi, T.; Karras, T.; Laine, S.; Lehtinen, J.; and Aila, T. 2019. Improved Precision and Recall Metric for Assessing Generative Models. In *Advances in Neural Information Processing Systems*, volume 32.

Lin, J. 1991. Divergence Measures Based on the Shannon Entropy. *IEEE Transactions on Information Theory*, 37(1): 145–151.

Lloyd, S.; and Weedbrook, C. 2018. Quantum generative adversarial learning. *Physical review letters*, 121(4): 040502.

- Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; and Geiger, A. 2019. Occupancy Networks: Learning 3D Reconstruction in Function Space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4): 1–15.
- Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 165–174.
- Pérez-Salinas, A.; Cervera-Lierta, A.; Gil-Fuster, E.; and Latorre, J. I. 2020. Data re-uploading for a universal quantum classifier. *Quantum*, 4: 226.
- Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; and Chen, X. 2016. Improved Techniques for Training GANs. In *Advances in Neural Information Processing Systems*, volume 29.
- Schuld, M.; Sweke, R.; and Meyer, J. J. 2021. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3): 032430.
- Silver, D.; Ranjan, A.; Patel, T.; Gandhi, H.; Cutler, W.; and Tiwari, D. 2023. MosaiQ: Quantum Generative Adversarial Networks for Image Generation on NISQ Computers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 7007–7016.
- Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33: 7462–7473.
- Sobel, I.; and Feldman, G. 1968. An Isotropic 3x3 Image Gradient Operator. Presentation at Stanford Artificial Intelligence Project. Reprinted in 2014.
- Su, K.; Chen, M.; and Shlizerman, E. 2022. Inras: Implicit neural representation for audio scenes. *Advances in Neural Information Processing Systems*, 35: 8144–8158.
- Thomas, A. M.; Youel, H.; and Jose, S. T. 2025. VAE-QWGAN: addressing mode collapse in quantum GANs via autoencoding priors. *Quantum Machine Intelligence*, 7(2): 91.
- Tsang, S. L.; West, M. T.; Erfani, S. M.; and Usman, M. 2023. Hybrid quantum–classical generative adversarial network for high-resolution image generation. *IEEE Transactions on Quantum Engineering*, 4: 1–19.
- Vieloszynski, A.; Cherkaoui, S.; Ahmad, O.; Laprade, J.-F.; Nahman-Lévesque, O.; Aaraba, A.; and Wang, S. 2024. Latentqgan: A hybrid qgan with classical convolutional autoencoder. In *2024 IEEE 10th World Forum on Internet of Things (WF-IoT)*, 1–7. IEEE.
- Wang, S.; Theobalt, C.; and Golyanik, V. 2026. Quantum visual fields with neural amplitude encoding. *Advances in Neural Information Processing Systems*, 38: 158535–158559.
- Yang, X.; Zhou, R.; Jia, S.; Koh, D. E.; Goh, S. T.; Li, Y.; Chen, H.; and Xiong, F. 2026a. End-to-End QGAN-Based Image Synthesis via Neural Noise Encoding and Intensity Calibration. *arXiv preprint arXiv:2603.18554*.
- Yang, X.; Zhou, R.; Jia, S.; Li, Y.; Yan, J.; Long, Z.; Guo, W.; Xiong, F.; and Xu, W. 2026b. iHQQAN: A lightweight invertible hybrid quantum-classical generative adversarial networks for unsupervised image-to-image translation. *Expert Systems with Applications*, 296: 128865.
- Zhang, J.; Che, X.; Fan, Y.; Peng, S.; Chen, G.; Ma, Q.; and Hu, J. 2025. Denoising diffusion models with optimized quantum implicit neural networks for image generation. *Future Generation Computer Systems*, 173: 107875.
- Zhao, J.; Qiao, W.; Zhang, P.; and Gao, H. 2024. Quantum implicit neural representations. In *Proceedings of the 41st International Conference on Machine Learning, ICML’24*. JMLR.org.
- Zoufal, C.; Lucchi, A.; and Woerner, S. 2019. Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, 5(1): 103.

Appendix 1 Classical components

The proposed framework contains two classical components: a classical embedding network in the generator and a convolutional critic for adversarial training. The embedding network transforms each spatial coordinate and latent variable into the rotation parameters of the quantum circuit, thereby conditioning the quantum generator on both pixel location and stochastic variation. The critic follows the WGAN-GP formulation and maps an input image to a scalar score through a sequence of strided convolutional layers.

The classical component of the generator is a classical embedding network. Each pixel coordinate is encoded using a sinusoidal positional embedding with six frequency bands, producing a 26-dimensional coordinate feature. This feature is concatenated with a 10-dimensional latent vector and passed through a three-layer MLP with hidden width 256 and ReLU activations. The MLP outputs 12 rotation angles, corresponding to three rotation angles for each of the four feature qubits. The raw angles are squashed by a tanh nonlinearity and scaled by π before being passed to the quantum circuit.

The discriminator is a classical WGAN-GP critic composed of three stride-2 convolutional layers with 3×3 kernels and channel widths (32, 64, 128), each followed by a LeakyReLU activation with negative slope 0.2. The resulting feature map is flattened and mapped to a scalar critic score by a fully connected layer. All convolutional and dense kernels are initialized with Glorot uniform initialization, and biases are initialized to zero. The detailed architectures and initialization settings of the two components are summarized in Table 1.

Module	Component	Configuration
Generator embedding	Coordinate input	Two-dimensional pixel coordinate (x, y)
	Positional encoding	Sinusoidal encoding with six frequency bands
	Coordinate feature dimension	26
	Latent input dimension	10
	MLP architecture	Three layers, hidden width 256
	Activation	ReLU
	Output	12 rotation angles for four feature qubits
	Output transformation	$\pi \tanh(\cdot)$
Critic	Architecture	Three stride-2 convolutional layers
	Kernel size	3×3
	Channel widths	32, 64, and 128
	Activation	LeakyReLU with negative slope 0.2
	Output layer	Flatten followed by a scalar-valued linear layer
Initialization	Weights	Glorot uniform initialization
	Biases	Zero initialization

Table 1: Architectural details of the classical components used in the final configuration with $N_f = 4$ feature qubits and $r = 20$ re-uploading layers.

Appendix 2 Training Convergence Analysis

To evaluate the generation quality and convergence behavior during training, Figure 1 presents the FID curves of the jointly trained MNIST 0–9 model and the individual digit-specific models. Since the joint dataset and the single-digit datasets have different target distributions and feature statistics, their FID values are not directly comparable in magnitude. Therefore, we primarily focus on the evolution of FID throughout training rather than its absolute value. To reduce random fluctuations and better illustrate the overall convergence behavior, the curves are smoothed using a moving average with a window size of five evaluation points, while the original FID curves are retained as thinner, lighter background lines to visualize the underlying training fluctuations. As shown in Figure 1, the FID decreases steadily throughout training for both the single-class models and

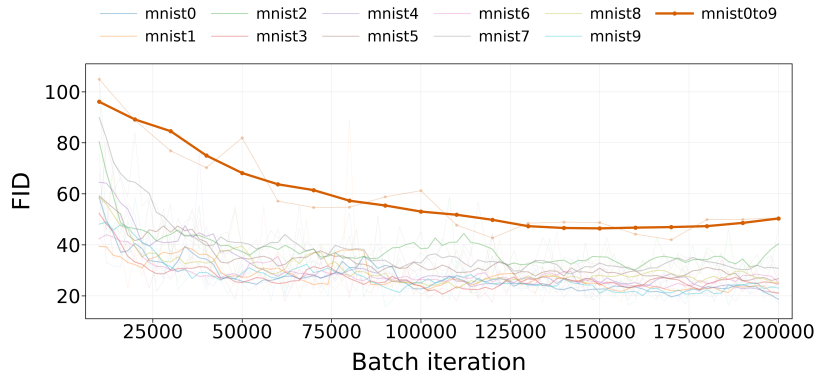


Figure 1: Training FID curves of the jointly trained MNIST 0–9 model and the individual digit-specific models.

the jointly trained multi-class model, and gradually stabilizes in the later stages. This indicates that the proposed model is able to progressively learn the target data distribution and achieve stable convergence under both single-class and multi-class settings.