Supplementary Material

This appendix supplements the main paper with expanded content, starting with a detailed background on gate-based quantum computing, covering its foundations in quantum physics and applications in machine learning and INRs (Sec. A). It explores practical implementations supported by QVF such as image inpainting (Sec. B) and shape completion (Sec. C), complemented by additional visualizations of 3D geometries (Sec. D) and the quantum ansatz architecture used in primary experiments (Sec. J). The appendix further investigates parameter scaling dynamics (Sec. E), outlines the full algorithmic protocol (Sec. F), and expands on related literature (Sec. G). Theoretical contributions include a proof demonstrating the efficacy of the ansatz in QVF, which feature bounded haar randomness, for enhanced learning (Sec. H), alongside an analysis linking quantum circuit measurements to Bayesian inference frameworks (Sec. I).

A Background

A.1 Preliminaries of Gate-based Quantum Computing

Qubits. The fundamental information blocks of a quantum processing unit (QPU) are qubits, i.e., the analogues of bits in classical computing. Unlike classical bits deterministically representing one possible state (0 or 1), qubits can statistically represent two distinct information states at the same time, denoted in the bra–ket notation as $|0\rangle$ and $|1\rangle$.

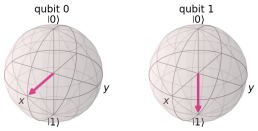

Figure 10: Bloch sphere visualisation of qubit states. Qubit 0: $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, qubit 1: $|\psi\rangle = |1\rangle$.

Superposition is a fundamental property distinguishing qubits from bits: It grants qubits the capacity to exist in a combinatorial state $|\psi\rangle$ of $|0\rangle$ and $|1\rangle$ such that:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{13}$$

with $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. Qubit states $|\psi\rangle$ can be visualised on Bloch spheres (see Fig. 10) or expressed in a vector form:

$$|0\rangle = \begin{bmatrix}1\\0\end{bmatrix}, |1\rangle = \begin{bmatrix}0\\1\end{bmatrix}, |\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle = \begin{bmatrix}\alpha\\\beta\end{bmatrix}. \tag{14}$$

Measurement in quantum mechanics inherently adopts a statistical approach to extract numerical information. For a qubit state $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ measured with operator $\hat{O}$ (that must be Hermitian, i.e., $\hat{O}^\dagger = \hat{O}$), this implies probabilities $|\alpha|^2$ and $|\beta|^2$, respectively, for measuring the information (i.e., eigenvalue of the measurement operator $\hat{O}$) stored in states $|0\rangle$ and $|1\rangle$:

$$\hat{O}\,|0\rangle = \kappa\,|0\rangle \text{ and } \hat{O}\,|1\rangle = \delta\,|1\rangle\,, \tag{15}$$

where κ and δ are eigenvalues of the measurement operator $|O\rangle$. The key aspect of measurement is the phenomenon known as wave function collapse, i.e., the projective measurement causes $|\psi\rangle$ to collapse to the operator's eigenstate, $|0\rangle$ or $|1\rangle$, conditioned on the measurement, i.e., κ or δ.

Entanglement further distinguishes quantum from classical computing. In the classical case, information stored in bits is independent, i.e., measuring one bit does not affect others. In the quantum realm, qubits can be highly correlated, exhibiting entanglement such that the information of one qubit can be interrelated with another, despite possible physical distance between them. For instance, a general information state of a 2-qubit system $|\psi\rangle_2$ can be expressed as:

$$|\psi\rangle_2 = a\,|00\rangle + b\,|01\rangle + c\,|10\rangle + d\,|11\rangle\,, \tag{16}$$

with $a, b, c, d \in \mathbb{C}$ such that $|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1$. The 2-qubit system is considered entangled if $|\psi\rangle_2$ cannot be expressed as a tensor product of two qubits $|\psi\rangle_{a1}$ and $|\psi\rangle_{a2}$, indicating that their information cannot be independently measured without disturbing each other, i.e.,

$$|\psi\rangle_2 \neq |\psi\rangle_{a1} \otimes |\psi\rangle_{a2}\,. \tag{17}$$

Rotation Operators. The operators responsible for rotating quantum states $|\psi\rangle$ of qubits along x, y, z axes on a Bloch sphere are referred to as rotation operators. Any single qubit operator $\hat{R}$ can be expressed as a combination of such rotation operators $\hat{R}_x, \hat{R}_y, \hat{R}_z$, i.e., $\hat{R}(\theta, \tau, \gamma) = \hat{R}_x(\theta)\hat{R}_y(\tau)\hat{R}_z(\gamma)$ with angles θ, τ and γ:

$$\hat{R}_x(\theta) = \begin{bmatrix}\cos(\frac{\theta}{2}) & -i\sin(\frac{\theta}{2})\\ -i\sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2})\end{bmatrix}, \quad \hat{R}_y(\tau) = \begin{bmatrix}\cos(\frac{\tau}{2}) & -\sin(\frac{\tau}{2})\\ \sin(\frac{\tau}{2}) & \cos(\frac{\tau}{2})\end{bmatrix}, \quad \hat{R}_z(\gamma) = \begin{bmatrix}e^{-i\frac{\gamma}{2}} & 0\\ 0 & e^{i\frac{\gamma}{2}}\end{bmatrix}. \tag{18}$$

The Pauli operators $\hat{X}, \hat{Y}, \hat{Z}$ represent specific instances of above rotation operators, i.e. rotations by π radians along the x, y, z-axes, respectively. These operators can also be expressed as matrices in the computational basis $|0\rangle, |1\rangle$ as follows:

$$\hat{X} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \hat{Y} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \hat{Z} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{19}$$

The Schrödinger Equation. Quantum computing involves the manipulation of information according to the principles of quantum mechanics, with its foundation rooted in the time-dependent Schrödinger equation:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H}(t) |\psi(t)\rangle, \tag{20}$$

where $\hbar$ is Planck's constant, and $|\psi(t)\rangle$ and $|\psi(0)\rangle$ are the quantum states after and before evolution, respectively; $\hat{H}$ is the Hamiltonian operator of the quantum system. Therefore, the evolution of quantum states can be described by the following relationship:

$$|\psi(t)\rangle = \hat{T} e^{-\frac{i}{\hbar}\int_0^t \hat{H}(t)dt} |\psi(0)\rangle, \tag{21}$$

with $\hat{T}$ denoting the time ordering operator. This simplifies to $e^{-\frac{it}{\hbar}\hat{H}} |\psi(0)\rangle$ for time-independent $\hat{H}$. Using a more compact notation, the Schrödinger equation can also be equivalently written as:

$$|\psi(t)\rangle = \hat{U}(\hat{H}, t) |\psi(0)\rangle, \text{ with} \tag{22}$$

$$\hat{U}(\hat{H}, t) = e^{-\frac{it}{\hbar}\hat{H}}. \tag{23}$$

To perform rotation operations on qubits, the system Hamiltonian $\hat{H}$ can be set to $E\hat{\sigma}$, with $\hat{\sigma} \in \{\hat{X}, \hat{Y}, \hat{Z}\}$. By substituting $\eta = 2Et/\hbar$, we arrive at:

$$\hat{U}(\hat{H}, t) = e^{-\frac{it}{\hbar}\hat{H}} = e^{-\frac{i\alpha}{2}\hat{\sigma}} = \hat{R_\sigma}(\eta). \tag{24}$$

A.2 Preliminaries of INR

For completeness, we provide a brief review of INR. INR models continuous signals—such as images, 3D geometries, or physical fields—as neural networks. Unlike traditional explicit representations that store data discretely (e.g., pixels or voxels), INR defines a signal as a continuous function $f_\theta : \mathbb{R}^d \to \mathbb{R}^k$, where a neural network parameterized by θ maps spatial or temporal coordinates to corresponding signal values, enabling resolution-independent, memory-efficient, and differentiable representations, making INR particularly valuable for tasks requiring interpolation, super-resolution, or solving inverse problems. The network is trained to minimize the reconstruction loss over sampled coordinates $\mathcal{L}(\theta) = \sum_{\mathbf{x}\in\mathcal{X}} \|f_\theta(\mathbf{x}) - \mathcal{S}(\mathbf{x})\|^2$ where $\mathcal{X}$ denotes the domain of interest. INR architectures typically address spectral bias—the tendency to favor low-frequency components—through positional encoding or periodic activation functions such as SIREN (45).

A.3 Quantum Machine Learning

The potential of quantum computing to enhance machine learning algorithms leads to the emergence of quantum machine learning (QML) (40), a discipline employing quantum mechanical phenomena to tackle classically intractable learning problems through enhanced computational paradigms (40). Central to QML are: 1) a feature map, which encodes classical input data into quantum states and 2) a variational ansatz, which performs quantum transformation on the quantum states. PQC have been shown to be asymptotic universal function approximators (5; 41). Several standardized QML algorithms have been explored, including quantum principal component analysis (28), quantum support vector machines (37), quantum Boltzmann machines (2), and quantum k-means clustering (24).

Feature Map. Integrating classical Euclidean data $\mathbf{x}$ into quantum computational frameworks necessitates a non-trivial mapping to quantum states $|\psi(\mathbf{x})\rangle$ in a Hilbert space $\mathcal{H}$. Several established encoding techniques exist, including *basis encoding*, *amplitude encoding*, *Hamiltonian evolution encoding*, with each presenting distinct trade-offs in qubit efficiency and circuit depth complexity. However, the determination of optimal encoding schemes remains an open research challenge, as the relationship between encoding fidelity $\mathcal{F}(\mathbf{x}) = |\langle\psi_{\text{ideal}}(\mathbf{x})|\psi_{\text{encoded}}(\mathbf{x})\rangle|^2$, resource requirements, and task-specific performance metrics (e.g., classification accuracy or function approximation error ϵ) remains poorly characterized across different problem domains.

Variational Ansatz. Quantum evolution of classical information embedded in states $|\psi(x)\rangle$ requires parameterized unitary ansatz $\hat{U}(\theta) \in \mathbb{C}^{2^n \times 2^n}$ acting on n-qubit systems. Physically, the ansatz is constructed through sequential composition of such unitary transformations, formally expressed as $\hat{U}(\boldsymbol{\theta}) = \mathcal{T}\left(\prod_{i=1}^t \hat{U}_i(\theta_i)\right)$, where $\mathcal{T}$ denotes the time-ordering operator governing gate sequence implementation. This induces a Hilbert

space transformation $\hat{U}(\boldsymbol{\theta}) : \mathcal{H} \to \mathcal{H}$ that maps input states to processed output states through the operation $|\phi(x, \boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})\,|\psi(x)\rangle$.

Measurement. Quantum computation culminates in statistical data extraction from evolved quantum states $|\phi(x)\rangle$ through projective measurements using Hermitian observables $\hat{O}$, where the computational output is formally defined as the expectation value: $V(x) = \langle\phi(x)|\,\hat{O}|\phi(x)\rangle$. Measurements collapse quantum states according to the Born rule, thereby restricting access to the embedded classical information to statistical estimators derived from repeated measurements. The choice of observable $\hat{O}$ fundamentally governs both the information-theoretic capacity of the measurement protocol and its computational complexity.

Training a Variational Ansatz. Instead of constructing a computational graph and performing backpropagation, training quantum circuits involves only forward evaluations (50). To minimize a measurement-dependent cost function $\mathcal{L}(\boldsymbol{\theta})$, the exact gradients $\nabla\mathcal{L}(\boldsymbol{\theta})$ can be evaluated through quantum circuit evaluations at shifted parameters $\boldsymbol{\theta} \pm \frac{\pi}{2}\mathbf{e}_i$ for basis vectors $\mathbf{e}_i$, expressed as:

$$\partial_{\theta_i}\mathcal{L} = \frac{1}{2}\left[\mathcal{L}\left(\theta_i + \frac{\pi}{2}\right) - \mathcal{L}\left(\theta_i - \frac{\pi}{2}\right)\right]. \tag{25}$$

This technique, i.e. parameter-shift rule, exploits the trigonometric structure of unitary gate generators $\hat{G}_i$ (where $\hat{U}_i(\theta_i) = e^{-i\theta_i\hat{G}_i}$) to enable hardware-compatible gradient estimation without numerical approximation or persistent circuit memory - a critical advantage over classical backpropagation that requires differentiable computational graphs.

A.4 Barren Pleataus

Training a variational ansatz $\hat{S}(\boldsymbol{\theta})$ is fundamentally constrained by the *barren plateau* phenomenon, where random parameter initialization induces exponential vanishing of cost function gradients across Hilbert space. As formally demonstrated by McClean et al. (30) through concentration of measure analysis:

"...for a wide class of reasonable parametrized quantum circuits, the probability that the gradient along any reasonable direction is non-zero to some fixed precision is exponentially small as a function of the number of qubits."

This observation is also known as *barren plateau*, which can be expressed mathematically for a system with n qubits as:

$$\mathbb{E}_w[\partial_w L(w)] = 0, \quad \mathrm{Var}_w[\partial_w L(w)] \in O(\frac{1}{\nu^n}),\ \nu > 1, \tag{26}$$

where ν characterizes the circuit's entangling capacity. The variance bound's scaling establishes that gradient estimators require $\mathcal{O}(\nu^n)$ measurement samples to maintain constant precision - an exponential resource overhead that renders practical optimization infeasible for $n \gg 1$. This poses challenges, particularly for gradient-based learning. Identified factors contributing to this phenomenon include observable locality (10; 46), specific noise models (48), and ansatz close to a 2-design, i.e., matching Haar random unitaries up to the second moment (30; 21). Those highlight the importance of selecting appropriate initialization protocols, quantum ansatz designs and observables.

B Application: Image Inpainting

QIREN (53) and 3D-QAE (36) are constrained by their reliance on fixed latent representations or rigid interpolation mechanisms, thereby not capable to reconstruct complete, coherent outputs from partial or corrupted inputs. QVF addresses such limitations by conditioning the quantum circuit topology on both the query point and a dynamic latent space, benefiting applications such as image inpainting. Given images with occluded or corrupted pixels, the circuit identifies in the latent space a vector that minimizes the discrepancy between the predicted multi-dimensional properties learned by the quantum circuit and the observed noisy values. The optimal latent vector conditioned quantum circuit enables recovery of missing field properties across the whole domain. Empirically, we masked out half of the image pixels and reconstructed the complete images via the protocol. Representative results of image inpainting with QVF are visualized in Fig. 11, demonstrating that QVF can deliver promising performance and accurately recover images even under such extreme sparsity, positioning QVF as a promising quantum circuit architecture for such tasks.

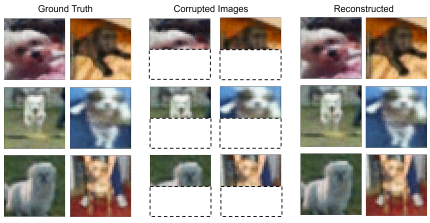

Figure 11: Image inpainting results with QVF.

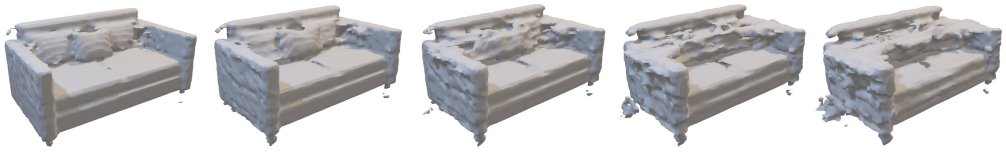

Figure 12: Shape completion from partial and noisy input depth maps using QVF; α is the noise ratio.

C Application: Shape Completion from Partial and Noisy Depth Maps

With the principle similar to image inpainting, QVF can perform tasks such as completing 3D geometries while existing quantum models can not. We adopt similar empirical settings as for images by cropping half of the samples along the depth (see Fig. 8-(b)) and consider the effects of zero-mean Gaussian noise across different perturbation ratios α to the depth map. Shape completion performance is quantified across incremental perturbation ratios, parameterized as $\alpha \in \{0, 0.005, 0.01, 0.02, 0.03\}$, where $\alpha = 0$ corresponds to an idealized noise-free completion scenario. Quantitative analysis reveals a monotonic decline in reconstruction fidelity with increasing noise perturbations, as evidenced by progressive geometric distortions and surface irregularities in Fig. 12.

D Additional 3D Geometry Visualizations

In conjunction with the quantitative results summarized in Tab. 3, Fig. 13-(a) provides a qualitative comparison of the reconstructed 3D geometries, contrasting the baseline—the classical architectural component in our model—with QVF. The baseline's numerically elevated loss values correlate with visual structural discontinuities, exemplified by the fragmented sofa leg, underscoring its propensity for topological inconsistencies during reconstruction. QVF, instead, demonstrates enhanced structural coherence, generating topologically intact geometries devoid of visible artifacts, as evidenced by its preservation of fine-grained features. Further geometric analysis, illustrated via color-encoded per-surface Hausdorff distance distributions in Fig. 13-(b), reveals systematic geometric deviations for the baseline (top) and ours (bottom), corroborating its geometric fidelity. An interesting observation is that the color distributions between models align well, meaning that QVF inherits the representative expressiveness of the classical component but enhances it due to inherent spectral connections to the quantum circuit.

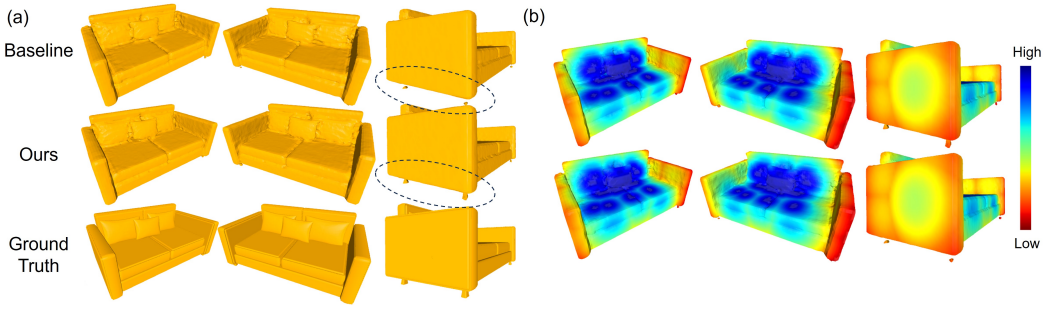

Figure 13: (a) Comparison of geometric representations using QVF and the classical model; ground truth is presented at the bottom; (b) representation fidelity visualized via color-encoded Hausdorff distance map: colors represent the distance to the ground truth. The rendered image employs a color gradient (blue>green>yellow>red) to indicate descending Hausdorff distance levels.

E Parameter Scaling Analysis

Complementing the visualization of model parameters in Fig. 9, we present a rigorous mathematical analysis of parameter scaling in QVF. The classical component, which we model as a dense neural network for inferring quantum system energy spectra, exhibits quadratic parameter growth $\mathcal{O}(p^2)$ with respect to hidden dimension p - a direct consequence of its fully connected architecture requiring pairwise neuron interactions. In contrast, the quantum evolution parameters demonstrate parametrically efficient linear scaling $\mathcal{O}(n\mathcal{J})$ for an n-qubit circuit of depth $\mathcal{J}$. This fundamental interdependence necessitates balanced co-design: while quantum circuit

scaling enhances expressivity through Hilbert space exploration, the operational fidelity critically depends on the classical network's capacity to provide accurate spectra. Optimal system performance requires concurrent optimization of classical representation capacity and quantum resource allocation.

F Algorithmic Protocol

We provide the complete training protocol of QVF in Alg. 1.

Algorithm 1 QVF Training Protocol

Input: Training dataset $\mathcal{D} = \{(x_i, s_i)\}_{i=1}^{N}$; number of qubits n; epochs N_{epoch}; measurement shots N_{shot}; parameters $\boldsymbol{\theta}_q$, $\boldsymbol{\theta}_c$; inverse temperature β.
for epoch $= 1$ **to** N_{epoch} **do**
Classical Inference:
- Compute energy spectrum $\mathbf{E}(\mathbf{x}; \boldsymbol{\theta}_c)$ via classical model.
- Evaluate Gibbs distribution:

$$P_i = \frac{e^{-\beta E_i}}{Z}, \text{ where } Z = \sum_{j=1}^{2^n} e^{-\beta E_j}.$$

Quantum State Preparation:
Initialize $\hat{\rho}_0 = \sum_{i=1}^{2^n} P_i \left|i\right\rangle \left\langle i\right|$.
Quantum Evolution:
Apply ansatz $\hat{S}(\boldsymbol{\theta}_q) = \prod_{\ell=1}^{J} e^{-i\theta_{q,\ell}\hat{H}_\ell}$ to obtain:

$$\hat{\rho}(\boldsymbol{\theta}_q) = \hat{S}(\boldsymbol{\theta}_q)\hat{\rho}_0\hat{S}^\dagger(\boldsymbol{\theta}_q)$$

Measurement & Observables:
Estimate $\langle \hat{O}_k \rangle = \text{Tr}[\hat{\rho}(\boldsymbol{\theta}_q)\hat{O}_k]$ for $k = 1, \ldots, K$.
Gradient Computation:
- Quantum: $\partial\langle \hat{O}_k \rangle / \partial\theta_{q,\ell}$ via parameter-shift rule.
- Classical: $\nabla_{\boldsymbol{\theta}_c}\mathcal{L}$ via chain rule through $\mathbf{E}(\mathbf{x}; \boldsymbol{\theta}_c)$.

Parameter Update:
Adam optimizer step with learning rate η:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \left(\nabla_{\boldsymbol{\theta}_c}\mathcal{L}, \nabla_{\boldsymbol{\theta}_q}\mathcal{L}\right).$$

end for
Output: Optimized parameters $\boldsymbol{\theta}_q^*$, $\boldsymbol{\theta}_c^*$.

G Extended Literature Discussion

Recent frameworks, including QIREN (53) and related works (42; 1; 27; 20; 23), adopt a unified architectural paradigm—termed the sandwich structure—in which a quantum circuit is embedded between classical preprocessing and postprocessing layers. Such a framework positions PQCs as trainable Fourier feature encoders, theoretically capable of projecting classical inputs into high-dimensional spectral representations under optimal conditions, subsequently processed by classical neural networks. While this design draws conceptual parallels to classical Fourier Feature Networks (FFNs) by substituting hidden layers with quantum circuits to explore quantum-enhanced feature spaces, it introduces fundamental conceptual and operational trade-offs that need to be carefully considered:

Over-reliance on Post-processing. The quantum circuit is constrained solely to be a feature generator, necessitating extensive post-processing by classical neural networks to extract meaningful information from the generated features. This dependency challenges claims of inherent quantum advantage, as the quantum subsystem does not directly contribute to the final inference step.

Blurring the Divide: Quantum Probabilism vs. Classical Determinism. Inference from a quantum circuit is naturally probabilistic by aggregating statistical projective measurements. Sandwiched structures, however, do not consider such a probabilistic nature but necessitate single-shot estimations for the inference of the quantum circuit before classical post-processing. This collapses the quantum subsystem's probabilistic nature into a

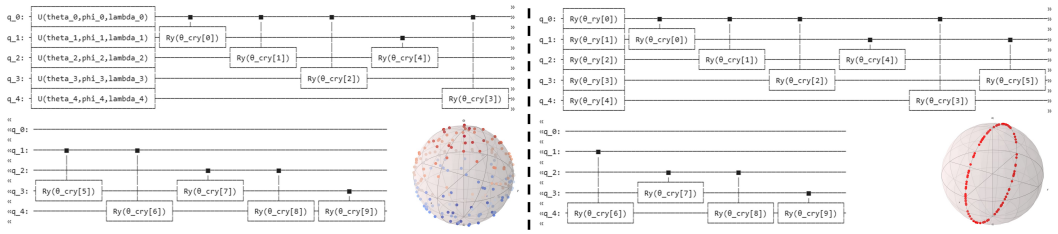

Figure 15: Visualization of ansatz designs and their exemplary induced traversable quantum states within Hilbert space: strongly entangled ansatz (left) and QVF (right). Traversable states of different circuit ansatz are visualized on the bottom right for both ansaetze.

deterministic feature map, as the final output becomes a parametric function of estimated expectation values rather than a direct manifestation of quantum measurement statistics.

H Variational Ansatz with Bounded Haar Randomness Improves Learning

We explain in more detail how our variational ansatz design featuring bounded Haar randomness could better preserve the gradient, which is crucial for inferring precise energy spectra. As a review, for quantum circuits with a generic parameterized ansatz, the gradient expectation value asymptotically scales to zero as the qubit count increases, a consequence of concentration in high-dimensional Hilbert spaces. The vanishing gradient phenomenon, i.e. barren plateaus, is thus governed by the variance decay rate, which determines optimization tractability. We quantify this phenomenon by analyzing the gradient variance $\mathrm{Var}_{\mathrm{grad}}$:

$$\mathrm{Var}_{\mathrm{grad}} = E_{\theta\sim U([0,2\pi))}\left[\mathrm{Var}\left(\left\{\frac{\partial}{\partial\theta_k}\langle\hat{H}\rangle_\theta\right\}_{k=1}^{M}\right)\right]. \tag{27}$$

We empirically evaluate our variational ansatz across an increasing number of qubits and benchmark against the established strongly entangled ansatz initially proposed in (39), and later adopted in QIREN. The results are shown in Fig. 14, confirming our ansatz design better preserves the gradients — a critical factor for our QVF design. Such behavior, as demonstrated by McClean et al. (30), can also be reflected via a higher KL divergence between the fidelity distribution of the variational ansatz and the Haar-random ensemble; see Tab. 4. Specifically, when the traversable quantum states of the variational ansatz closely resemble Haar-random states, its parameter landscape becomes flat, leading to vanishing gradients and poor optimization. Concurrently, bounding the Haar randomness of the ansatz would not impact the expressivity of the circuit, as Z-basis measurement probabilities remain invariant to the real-subspace constraint, as they solely depend on magnitudes of state amplitudes. We evaluate how the trainability disparity contributes to performance on learning image representations. Quantification results are recorded in Tab. 4.

	Baseline (39)	Ours
KL divergence	0.0491	0.2392
MSE ($1e^{-3}$)↓	1.66	**1.02**

Table 4: Performance comparison.

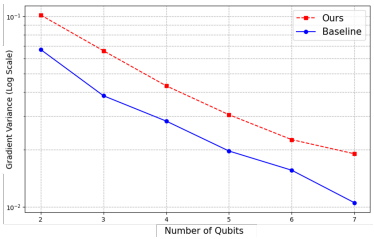

Figure 14: Gradient variance across circuit of different scales.

I Connection of PQCs to Bayesian Inference

As quantum circuits are inherently probabilistic models, they share conceptual parallels with Bayesian inference frameworks. In Bayesian neural networks (BNNs), probabilistic outputs emerge from parameters governed by prior distributions, with training focused on maximizing the conditional likelihood of observed data labels while implicitly updating a posterior distribution over the parameters. For PQC, while they leverage deterministic parameters, they exhibit probabilistic outputs due to the stochastic nature of quantum measurements, which—under sufficiently large shot counts—approximate Gaussian distributions in accordance with the CLT. While the probabilistic outputs of PQCs permit an interpretative lens rooted in Bayesian principles, their training does not inherently involve posterior inference over parameters unless explicitly cast within a Bayesian formalism. This distinction underscores that the Bayesian interpretation of PQCs arises from their measurement statistics rather than an intrinsic probabilistic parameter space.

J Detailed Variational Ansatz Visualization

We provide a visualization of: 1) the QVF ansatz and 2) the widely-adapted strongly-entangled ansatz (39) in Fig. 15. We highlight the circuit structures via their parametrized single-qubit rotations and inter-qubit

entanglement patterns. Traversable quantum states are visualized on the Bloch sphere by sampling the ansatz. As expected, our circuit design confines generated quantum states to a circular subset of the Bloch sphere, whereas quantum states evolved by the entangling ansatz are broadly distributed across the complete Bloch sphere surface. Such visualizations complement our observation in Tab. 4, where the ansatz in QVF achieves an elevated KL divergence with Haar random states, leading to stronger gradient magnitudes and improved trainability.