

---

# Quantum Visual Fields with Neural Amplitude Encoding

---

Shuteng Wang      Christian Theobalt      Vladislav Golyanik\*  
MPI for Informatics, SIC, Germany

## Abstract

Quantum Implicit Neural Representations (QINRs) have emerged as a promising paradigm that leverages parametrised quantum circuits to encode and process classical information. However, significant challenges remain in areas such as ansatz architecture design, the effective utility of quantum-mechanical properties, training efficiency, and the integration with classical modules. This paper advances the field by introducing a novel QINR architecture for 2D image and 3D geometric field learning, which we collectively refer to as Quantum Visual Field (QVF). QVF encodes classical data into quantum statevectors using neural amplitude encoding grounded in a learnable energy manifold, ensuring meaningful Hilbert space embeddings. Our ansatz follows a fully entangled design of learnable parametrised quantum circuits, with quantum (unitary) operations performed in the real Hilbert space, resulting in numerically stable training with fast convergence. QVF does not rely on classical post-processing—in contrast to the previous QINR learning approach—and directly employs measurements to extract learned signals encoded in the ansatz. Experiments on a quantum hardware simulator demonstrate that QVF outperforms existing quantum approach and competes widely used classical foundational baselines in terms of visual representation accuracy across various metrics and model characteristics. We also show applications of QVF in 2D and 3D field completion and 3D shape interpolation, highlighting its practical potential. Project page: <https://4dqv.mpi-inf.mpg.de/QVF/>.

## 1 Introduction

Implicit neural representations (INRs) have emerged as a powerful framework for continuously modelling signals via neural networks. They are widely used in image and 3D shape synthesis, as well as 3D reconstruction, among other fields of visual computing [54]. INRs map spatial (and also temporal) coordinates to corresponding signal values, enabling resolution-independent, memory-efficient, and differentiable representations; the signal encoding network  $f_\theta$  with parameters  $\theta$  is trained to minimise the reconstruction loss  $\mathcal{L}(\theta)$  over sampled coordinates  $\mathbf{x}$ :  $\mathcal{L}(\theta) = \sum_{\mathbf{x} \in \mathcal{X}} \|f_\theta(\mathbf{x}) - \mathcal{S}(\mathbf{x})\|^2$ , where  $\mathcal{X}$  denotes the sampled domain and  $\mathcal{S}(\mathbf{x})$  is the signal value to be represented by  $f_\theta$ . As a remedy to the growing computational, memory and energy demand required by INR, recent work has explored the integration of quantum circuits into INR as a promising alternative to classical methods, with potential advantages in model compactness and learning efficiency [57]. Quantum algorithms operate within Hilbert spaces, enabling superposition and entanglement of states that facilitate parallel processing beyond classical systems with comparable resource scales. Specifically, quantum machine learning (QML) models involving parameterised quantum circuits (PQCs) or *ansatz* parameterise the evolution of quantum states through unitary transformations (implemented as quantum gate sequences), requiring a number of parameters that scale logarithmically with the Hilbert

---

\*Corresponding author’s email: [golyanik@mpi-inf.mpg.de](mailto:golyanik@mpi-inf.mpg.de)

space dimensionality. Recent studies further reveal an intrinsic link between PQCs and Fourier-based learning mechanisms [43], a critical feature for relieving biased learning or mitigating spectral bias common in classical neural networks for INRs [35]. Together, these insights suggest a pathway towards highly efficient and expressive QML models for visual computing.

Despite their theoretical promise, quantum implicit neural representations (QINRs) remain heavily underexplored. The recently introduced QIREN approach [57] is, to our knowledge, among the first in the field that is closest to our work, designed for image representation, upsampling and generation. In detail, QIREN projects query coordinates into learnable Fourier features paired with a classical network decoder, explicitly linking it to Fourier neural architectures while, at the same time, overshadow the quantum behaviour due to classical post-processing.

In response to these limitations, we propose Quantum Visual Field (QVF), a novel coordinate-based QML model that leverages high-dimensional Hilbert spaces for lightweight and spectrally unbiased implicit visual field representations; see the scheme in Fig. 1-(a). Rather than using heuristic classical-to-quantum data encoding methods [52, 39, 8, 36]—which (unreasonably) assume that handcrafted embeddings align with the inductive biases of quantum circuits—we introduce a novel learnable energy module that encodes non-linear data priors to align classical Euclidean and quantum Hilbert feature representations. Our carefully designed quantum circuit leverages quantum state evolution and entanglement between qubits to effectively explore the optimal pre-measurement quantum state representations. Specifically, the reachable Hilbert space is constrained in QVF for stable gradient flow and relief of issues such as barren plateaus, i.e., vanishing gradients arising due to Haar randomness<sup>2</sup>, without compromising expressiveness. The quantum circuit is measured to generate multi-dimensional signals, such as images or 3D geometries, or their collections (Fig. 1-(b)) through conditioning on latent variables. In summary, the technical contributions of this paper include:

- QVF, a coordinate-based QML model for visual representation learning (2D images and 3D signed distance fields). The QVF approach is designed for execution on quantum machine simulators or fault-tolerant gate-based quantum computers.
- A non-linear neural scheme for encoding classical data into quantum statevectors. Our neural amplitude encoding is grounded in a learnable energy manifold ensuring meaningful Hilbert space embeddings.
- An efficient PQC that processes entangled information within the real Hilbert subspace, explicitly designed for stable gradient feedback by bounding Haar randomness.

Unlike existing approaches [57], QVF is a lightweight architecture with the exact structural configuration dynamically depending on the input data. We evaluate QVF and compare it to the main competitor, i.e., prior QINR method QIREN (on 2D image representation learning), and, additionally, several foundational classical INR baselines (for 2D image and 3D shape representation learning); experiments are performed on a high-end simulator of gate-based quantum hardware [6]. We show that QVF consistently competes and outperforms QIREN and other compared techniques. Moreover, QVF supports problem scales and applications beyond the reach of prior QINR frameworks, such as image inpainting, shape completion and latent space interpolation (Fig. 1-(c)), taking a step towards unlocking quantum models in real scenarios.

<sup>2</sup>Haar randomness refers to the property of sampling quantum states uniformly at random from the Hilbert space according to the Haar measure. [19]

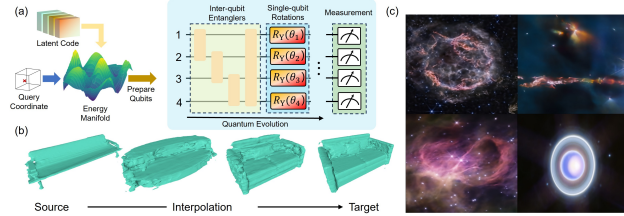


Figure 1: **Our learnable coordinate-based QVF model can represent various visual fields:** (a) Schematic diagram of the architecture; (b) Latent space interpolation of 3D signed distance fields [10]. (c) 2D image representation of a moderate resolution (400×350 pixels) [15];

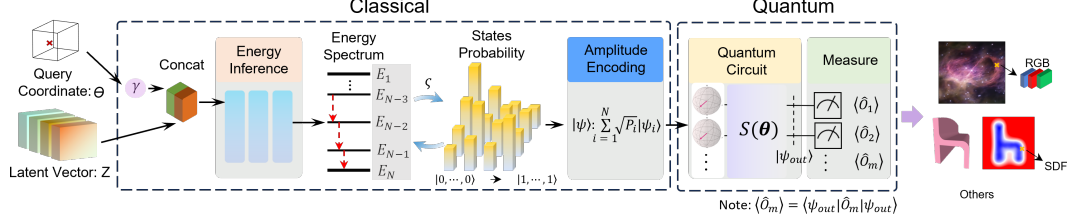


Figure 2: **Overview of the proposed QVF model**, a QML framework for visual representation learning. Query coordinates  $\Theta$  encoded using  $\gamma$  (positional encoding) concatenated with the conditioning latent code  $z$  are used to infer the energy spectrum  $E$  of a quantum system, associated with Boltzmann-regulated statistical uncertainty  $P$ . The inferred statistical property is leveraged in encoding the classical data into quantum statevectors, which are subsequently processed by the parametrised quantum circuit  $S(\theta)$ . Field properties are decoded probabilistically from projective circuit measurements.

## 2 Related Work

**Classical Neural 2D/3D Scene Representation.** Neural networks serve as a basis for modern implicit scene representations, employing continuous function approximations to circumvent the constraints of discrete grid-based methodologies [33, 11, 50, 12, 25]. Initial breakthroughs utilised multi-layer perceptrons (MLPs) to establish coordinate-to-attribute mappings, as demonstrated in Chen et al.’s [12] continuous implicit model for arbitrary-scale super-resolution. The paradigm has since been extended to 3D representations, supplanting conventional voxel- and mesh-based approaches: DeepSDF [34] achieves geometrically coherent surface reconstruction via learned signed distance fields (SDFs), while neural radiance fields (NeRF) [31] introduce a volumetric scene representation parameterized by coordinate-based neural mappings of spatial coordinates and viewing directions to radiance and density, enabling photorealistic novel view synthesis. Gate-based quantum computing offers great potential for fundamentally enhancing INRs.

**Quantum-enhanced Computer Vision.** Growing interest in quantum computing for computer vision has established quantum-enhanced computer vision (QeCV) as an emerging research frontier. Current literature predominantly explores quantum annealers for combinatorial optimisation [16, 7, 55, 14, 5, 29], while tunable quantum circuits remain underexplored. Early works introduced foundational concepts such as quantum image denoising via localized convolutional operations [44] and quantum convolutional neural networks (QCNNs) with mid-circuit measurements to emulate translational equivariance [13]. Subsequent advances include hybrid quantum-classical architectures for 3D point cloud classification through voxelization and quantum feature processing [2], as well as quantum autoencoders for classical data compression via hand-crafted amplitude embeddings [36].

Our work is inspired by 3D-QAE [36], who developed hand-crafted quantum amplitude embeddings for encoding 3D point clouds. However, their method, as acknowledged by their authors, suffers from limited scalability and underperforms classical models. Another related work is QIREN by Zhao et al. [57], which employs the sandwich structure, i.e., a quantum circuit layer placed between classical pre- and post-processing. They leverage circuit’s Fourier connections to project queries in the Fourier basis, followed by a classical dense layer for inference. This draws parallels to classical positional or Fourier encodings, with the Fourier spectrum size growing exponentially. While theoretically motivated, its practical utility is debated as heavy classical postprocessing reduces the quantum component to a feature generator. In contrast, our framework avoids such post-processing and more heavily relies on the ansatz; see an algorithmic comparison in Table 1. We use a learnable, energy-based Boltzmann-regulated amplitude encoding, which is a critical step towards unlocking the potential of quantum computing as demonstrated empirically in Sec. 5. At the same time, our carefully designed ansatz evolves the encoded data and ensures robust gradient feedback.

| Characteristic           | Ours        | QIREN [57]     | 3D-QAE [36]       |
|--------------------------|-------------|----------------|-------------------|
| No heavy post-processing | ✓           | ✗              | ✓                 |
| Data encoding            | Neural AE   | Neural Angular | AE (hand-crafted) |
| Qubit budget             | logarithmic | linear         | logarithmic       |
| Supported dimensions     | 2D/3D       | 2D             | 3D                |
| Quantum hardware         | Simulator   | Simulator      | Simulator         |

Table 1: Comparative algorithmic analysis of related work. “AE” denotes amplitude encoding.

### 3 Review: QML, its Unitary Nature and Fourier Structure

This paper assumes familiarity with quantum computing and its notations. For convenience, we provide a refresher in App. A. QML leverages parametrized unitary quantum operations on encoded data  $|\psi(x)\rangle = \sum_j \psi_j(x)|j\rangle$  to learn functions typically expressed as expectation values  $f(x) = \langle\psi(x)|\hat{O}|\psi(x)\rangle$ , where  $\hat{O}$  is a Hermitian observable ( $\hat{O} = \hat{O}^\dagger$ ). The unitary nature of quantum evolution ( $U^\dagger U = UU^\dagger = I$ ) preserves inner products and norms, ensuring that the spectral components of the encoded data are transformed by quantum circuits. The spectral decomposition of  $\hat{O} = \sum_k \lambda_k |e_k\rangle\langle e_k|$  reveals a Fourier-like structure in  $f(x) = \sum_k \lambda_k |\langle e_k|\psi(x)\rangle|^2$ , where the projections  $\langle e_k|\psi(x)\rangle$  act as Fourier coefficients and the eigenvalues  $\lambda_k$  relate to accessible frequencies. This fundamental property is a consequence of quantum mechanics: unitary transformations preserve spectral components, ensuring that even complex quantum circuits inherently operate in a frequency domain. As a result, the expressivity of QML models is directly linked to their accessible frequency components, influencing their ability to generalize and learn structured data representations.

### 4 Our QVF Approach

This section introduces the proposed QVF model, i.e., a QINR for learning visual representations and their collections; see Fig. 2 for its architecture. QVF takes query coordinates  $\Theta$  and an optional latent variable  $z$  (in the case more than one visual field needs to be represented) as inputs and produces 2D or 3D field properties  $s$ . We introduce encoding classical data into quantum states using neural amplitude encoding in Sec. 4.1, while the quantum circuit design and measurement are detailed in Sec. 4.2. Sec. 4.3 provides training details and applications supported by QVF.

#### 4.1 Amplitude Encoding with Neural Embeddings

Our parametrised energy-based embedding of classical data  $x$  into quantum states  $|\psi_{in}(x)\rangle$  generalises widely used hand-crafted amplitude encoding (AE) [36, 40]. AE enables an exponentially compact encoding of  $N = 2^n$  classical values into probability amplitudes of  $n$  qubits by leveraging quantum superposition. Notably, this implies that AE induces *exponentially-many* random Fourier features due to the inherent periodicity of quantum state phases. The fundamental limitation of hand-crafted AE stems from its *a priori, possibly biased* prepared quantum states, which poses a risk of misalignment with subsequent quantum evolution or suboptimal utilisation of task-specific data Fourier priors. Hence, we propose a data-driven approach for AE that learns the optimal quantum state density  $\hat{\rho}_{\text{opt}}(x)$ , directly from data, i.e., for QINR in our case. We restrict the process on pure quantum states satisfying  $\text{Tr}(\hat{\rho}(x)^2) = 1$ . Drawing upon the fundamental energy-probability duality inherent in physical systems (e.g., in statistical and quantum mechanics), we infer the conditional energy spectrum  $\mathbf{E}$  of a given visual representation and transform it into a probability distribution  $\mathbf{P}$  subsequently encoded as qubit state amplitudes  $\alpha_i \in \mathbb{C}$  residing in the complex Hilbert space  $\mathcal{H}$ . Our encoding introduces *non-linearity into the quantum evolution* while reserving the full repertoire of quantum processing and measurements<sup>3</sup>. For energy inference, we employ a minimal dense MLP  $f(x = \{\Theta, \mathbf{z}\}) : \gamma(\Theta) \times \mathbf{z} \rightarrow \mathbf{E}$  activated by ReLU; we use positional encoding to accelerate learning [35, 31]:

$$\gamma(\Theta) = (\dots, \sin(2^{L-1}\pi\Theta), \cos(2^{L-1}\pi\Theta), \dots). \quad (1)$$

$\Theta$  denotes the field query coordinate while  $\mathbf{z}$  represents the latent code in the case of learning visual field collections. The inferred  $\mathbf{E} = f(\gamma(\Theta), \mathbf{z})$  is leveraged to derive the Boltzmann-regulated  $\mathbf{P}$  of the quantum system; Gibbs-Boltzmann framework  $\varsigma$  serves as an inductive embedding bias for encapsulating thermodynamic uncertainty, enabling the realisation of Gibbs quantum states [1, 3]. We next formulate  $\mathbf{P} = [P_i], i \in \{1, \dots, N\}$  through the construction of a discretised energy landscape  $\mathbf{E}$ , derived from the Gibbs canonical ensemble:

$$\mathbf{P} = \frac{\exp(-\beta \mathbf{E}(\Theta, \mathbf{z}))}{\mathcal{Z}}, \text{ where} \quad (2)$$

$$\mathcal{Z} = \int \exp(-\beta \mathbf{E}(\Theta, \mathbf{z})) d\Theta \approx \sum_{j=1}^N \exp(-\beta E_j(\Theta, \mathbf{z})), \quad (3)$$

<sup>3</sup>transformations in quantum circuits before measurement are linear, which is often seen as a key limitation in ansatz expressivity

with  $\beta = (k_B T)^{-1}$  representing the inverse temperature. The quantum state amplitudes  $\alpha_i \in \mathbb{C}$  residing in the complex Hilbert space  $\mathcal{H}$  are characterised by their complex phases  $\phi_i = \arg(\alpha_i)$  which are arbitrary within the interval  $[0, 2\pi)$ ; it is subjected to the normalisation condition  $\|\alpha_i\|_2^2 = P_i$  with  $\sum_{i=1}^N P_i = 1$ . Finally, the quantum states  $|\psi_{in}(\Theta, \mathbf{z})\rangle$  encoding the query coordinates for our input fields are prepared as follows:

$$|\psi_{in}(\Theta, \mathbf{z})\rangle = \sum_{i=1}^N \alpha_i |\psi_i\rangle, \quad \alpha_i = \sqrt{P_i} e^{i \cdot \arg(\alpha_i)},$$

$$\hat{\rho}(\Theta, \mathbf{z}) = |\psi_{in}(\Theta, \mathbf{z})\rangle \langle \psi_{in}(\Theta, \mathbf{z})| = \sum_{i,j=1}^N \alpha_i \alpha_j^\dagger |\psi_i\rangle \langle \psi_j|. \quad (4)$$

$|\psi_i\rangle$  is the computational basis,  $\hat{\rho}(\Theta, \mathbf{z})$  is the density distribution of  $|\psi_{in}(\Theta, \mathbf{z})\rangle$  and “ $(\cdot)^\dagger$ ” denotes the adjoint. We then theoretically analyse data encoding effects on the model expressiveness.

**Lemma 1** *Energy inference exhibits functional equivalence to determining optimal non-linear input-dependent frequency spectrum embedded within variational quantum circuits, defining model’s inherent expressiveness.*

As demonstrated by Schuld et al. [43], variational circuits of the form  $U(x) = W^2 g(x) W^1$  admit a truncated Fourier-type expansion when measuring circuit expectation values  $\langle \hat{M} \rangle$ :

$$\langle \hat{M} \rangle = \langle 0 | W^{1\dagger} g(x)^\dagger \hat{M} g(x) W^1 | 0 \rangle = \sum_{w \in \Omega} c_w e^{i w x}, \quad (5)$$

where  $W^1$  and  $W^2$  are arbitrary unitary matrices. The effective measurement operator is defined as:  $\tilde{M} = W^{2\dagger} \hat{M} W^2$ , while  $g(x)$  serves as data encoding modules applied to the physical system. Notably, unlike Schuld et al. [43], where encoding gate analysis are restricted to Pauli gates,  $g(x)$  compass more general quantum operations. Prepared input quantum state  $|\psi_{in}(x)\rangle$  can be equivalently expressed as

$$|\psi_{in}(x)\rangle = g(x) W^1 |0\rangle, \quad (6)$$

establishing a direct correspondence between inferred energy landscape and the multi-dimensional frequency spectrum  $\Omega$ , with dependencies encoded in the learnable energy inference framework. The optimal effective measurement basis, given by  $\hat{M}_{\text{opt}} = W_{\text{opt}}^{2\dagger} \hat{M} W_{\text{opt}}^2$ , along with the heuristic learnable circuit design  $\hat{S}(\theta)$ , which approximates  $W_{\text{opt}}^2$ , will be introduced in the next sections.

## 4.2 Our Parametrised Quantum Circuit

Once the classical data is encoded into  $|\psi_{in}(\Theta, \mathbf{z})\rangle$ , it is processed by our learnable PQC or ansatz  $\hat{S}(\theta)$  within a high-dimensional Hilbert space. Our goal is a compact and expressive PQC for QINR learning. As unrestricted traversal of the Hilbert space can induce training instabilities, we therefore constrain the ansatz  $\hat{S}(\theta)$  to the manifold of real-valued unitaries, constructed from Pauli-Y rotations and entangling gates to ensure efficient training of QVF. This design choice avoids the imaginary components introduced by Pauli-X and -Z rotations, which would otherwise permit unconstrained exploration of the full, complex Hilbert space and hinder training. Our PQC architecture is analogous to classical densely-connected neural networks in the sense that it contains alternating layers of parametrised single-qubit Pauli-Y rotations and entangling operations, supporting highly correlated, non-local quantum states that cannot be decomposed into a tensor product of individual qubits which enables parallel information processing. This design provides critical benefits for QINR learning: (1) it confines state evolution to the manifold of real-valued unitaries, eliminating redundant parameter dimensions that facilitate scrambled quantum states and barren plateaus; (2) it naturally discards complex phase information while *preserving all measurement-relevant quantities*, i.e., *universality*, for certain basis observations such as Pauli-Z observable as the complex phase factors cancel out and become irrelevant when computing Z-basis probabilities; and (3) it maintains full expressivity while significantly simplifying the optimisation landscape. Additionally, our design necessitates

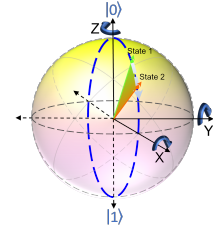


Figure 3: Representative pure qubit state transitions on the Bloch sphere.

the enforcement of a zero complex phase in our encoded quantum data, formally expressed as  $\arg(\alpha_i) = 0$ , with schematic Bloch sphere dynamics depicted in Fig. 3. Once  $|\psi_{in}(\Theta, \mathbf{z})\rangle$  has been transformed by  $\hat{S}(\theta)$ , we extract the visual field attributes encoded in our QINR using projective measurements of the final quantum states.

**Multi-dimensional Measurement.** To extract an  $m$ -dimensional representation ( $m \leq n$ ) from  $\hat{S}(\theta)$ , we implement local Pauli projective measurements on the first  $m$  qubits, effectively tracing out the remaining  $n-m$  qubits. The corresponding family of local measurement operators  $\{\hat{O}_i\}$  is formally defined over the  $n$ -qubit Hilbert space as:

$$\hat{O}_i = (\otimes_{k=1}^{i-1} \mathbb{I}_k) \otimes \sigma_i^Z \otimes (\otimes_{l=i+1}^n \mathbb{I}_l), \quad i \in \{1, \dots, m\}, \quad (7)$$

where  $\mathbb{I}_k$  denotes the identity operator acting on the  $k$ -th qubit, preserving its quantum state within the tensor product. The operator  $\sigma_i^Z = |0\rangle\langle 0|_i - |1\rangle\langle 1|_i$  represents the Pauli-Z observable applied to the  $i$ -th qubit. Local measurements help guarantee robust gradient feedback [9, 48] for the circuit. The output of  $\hat{S}(\theta)$  is defined as the expectation value of finite-shot circuit measurements (App. D analyses the influence of circuit measurements on the extracted image quality). This expectation value can be expressed as  $V_{\text{inf}}$ , which is defined as the number of shots approaches infinity in the asymptotic limit:

$$V_{\text{inf}}(\Theta) = \text{Tr}(\hat{\rho}(\Theta) \hat{M}(\theta)), \quad \hat{M}(\theta) = \hat{S}(\theta)^\dagger \hat{O} \hat{S}(\theta). \quad (8)$$

$\hat{M}(\theta)$  represents a parametrised measurement basis employed to approximate the optimal measurement basis  $\hat{M}_{\text{opt}}$  via unitary quantum evolutions of our fully-entangled circuit  $\hat{S}(\theta)$ . Similar to classical universal approximation theory, quantum circuits with sufficient depth can approximate arbitrary unitary transformations. The Solovay–Kitaev Theorem provides a rigorous theoretical upper bound on the number of quantum gates required to approximate an arbitrary unitary operation to a given precision  $\epsilon$ , given by  $O(4^n \log^4(1/\epsilon))$ . Circuit depth  $\mathcal{J}$ —analogous to the number of layers in classical neural networks—and the corresponding total number of constituent unitary quantum gates are hyperparameters of our  $\hat{S}(\theta)$ . The local measurement  $V(\Theta)_i$  corresponding to the  $i$ -th qubit is injectively mapped to the corresponding dimension of the target field, requiring the qubit number  $n \geq m$ . We next detail the end-to-end training protocol of our model under the Bayesian framework.

### 4.3 Training Details and Applications

**Initialisation of  $\hat{S}(\theta)$ .** We incorporate established PQC initialisation strategies, i.e., identity [18] and Gaussian [56]; see Fig. 4 for the architectural implications. For identity initialisation, each circuit layer  $\hat{S}(\theta)_j$  at depth  $j \in \{1, \dots, \mathcal{J}\}$  can be expressed as a sub-circuit  $S_R$  followed by its Hermitian adjoint  $F_R$ . Then,  $\hat{S}(\theta)_j$  is constructed by assigning  $\theta_k \sim \mathcal{U}[0, 2\pi)$  to  $S_R$ , which also initialised as  $F_R$  per definition. This ensures that the composite operation  $\hat{S}(\theta)_j = S_R F_R$  is equivalent to a zero circuit depth (identity circuit) before training. Note that while the initial configuration enforces  $S_R F_R = I$ , this constraint is not maintained during optimisation [18]. For the Gaussian initialisation, trainable parameters for  $\hat{S}(\theta)_j = S_{G_j}$  are, instead, sampled from a zero-mean Gaussian distribution with the variance coupled to the circuit depth, i.e.,  $\theta_k \sim \mathcal{N}(0, \sigma^2(\mathcal{J}))$  [56]. The overall circuit architecture  $\hat{S}(\theta)$  is obtained by concatenating  $\mathcal{J}$  blocks  $\hat{S}(\theta)_j$  such that the overall unitary transformation is given by  $\hat{S}(\theta) = \prod_{j=1}^{\mathcal{J}} \hat{S}(\theta)_j$ . Note that architectural homogeneity across blocks is maintained, preserving systematic exploration of the unitary space  $\mathcal{U}(2^n)$ .

**QVF Training.** Consider dataset  $X$  composed of  $W$  distinct visual fields denoted by  $X_i$ , for  $i \in \{1, \dots, W\}$ . Each data field  $X_i$  encapsulates physical field properties  $s_i^j$ , such as pixel values in images or SDF for geometric representations, sampled at specific spatial coordinates  $\Theta_i^j$ ; here, index  $j$  denotes the sample index per field. The relationship between spatial coordinates and physical properties is defined by a function  $f$ , such that sampled points within each field are given by:

$$X_i = \{(\Theta_i^j, s_i^j) | s_i^j = f(\Theta_i^j), j \in \{0, 1, \dots, M\}\}, \quad (9)$$

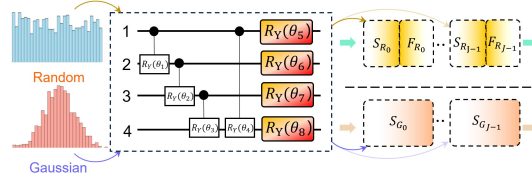


Figure 4: **QVF Initialisation:** Schematic circuit module initialised with identity (top) and Gaussian (bottom) schemes.

where  $M$  is the number of samples per field. Crucially, each data field  $X_i$  is associated with a unique latent code  $\mathbf{z}_i$ . The training objective is to maximise conditional probability distribution  $p_{\theta}(\mathbf{s}|\Theta)$ ;  $\theta$  represents trainable parameters of a QVF:

$$p_{\theta}(\mathbf{s}|\Theta) = \sum_{i,j} p_{\theta}(s_i^j|\Theta_i^j, \mathbf{z}_i) p(\mathbf{z}_i). \quad (10)$$

Under a sufficiently large number of i.i.d. quantum circuit measurement shots, the conditional likelihood  $p_{\theta}(s_i^j|\Theta_i^j, \mathbf{z}_i)$  can be approximated by a Gaussian distribution. This statement is supported by the Central Limit Theorem (CLT), which establishes the asymptotic normality of the sum (or average) of numerous i.i.d. random variables with finite variance. Consequently,  $p_{\theta}(s_i^j|\Theta_i^j, \mathbf{z}_i)$  can be approximated as

$$p_{\theta}(s_i^j|\Theta_i^j, \mathbf{z}_i) \approx \exp\left(-\mathcal{L}^2(V(\mathbf{z}_i, \Theta_i^j; \theta), s_i^j)\right), \quad (11)$$

where  $\mathcal{L}(\cdot)$  represents the loss function that quantifies the discrepancy between the output of the circuit  $V(\mathbf{z}_i, \Theta_i^j; \theta)$  and the observed physical property  $s_i^j$ . Model training can, therefore, be formulated as maximising this conditional likelihood under a Bayesian framework. To ensure a smooth representation transition in the latent space, the prior distribution over  $\mathbf{z}_i$  is softly penalised to follow a smooth distribution; an isotropic zero-mean multivariate Gaussian distribution is a reasonable choice as adopted by Park et al. [34]. The loss function  $L_{\theta, \mathbf{z}}$ , minimised via training over all learned  $W$  fields  $\{s_i|i = 1, \dots, W\}$  with  $M$  samples per  $s_i$ , is formulated as:

$$L_{\theta, \mathbf{z}}(\Theta, \mathbf{s}) = \sum_{i,j=1}^{W,M} (\mathcal{L}(V(\mathbf{z}_i, \Theta_i^j; \theta), s_i^j) + \gamma \|\mathbf{z}_i\|_2). \quad (12)$$

QVF undergoes end-to-end training: classical parameters are updated via gradient descent, while quantum parameters are optimised using the parameter-shift rule [32].

**Usage and Applications.** Once trained, we can query QVF for the encoded 2D or 3D representations in a coordinate-based manner. We can also infer with partial samples, enabling applications such as image inpainting and partial shape completion through latent space optimisation. Using Maximum-a-Posteriori (MAP) estimation, we identify a latent code  $\hat{\mathbf{z}}$  that maximises agreement with the input partial observation  $\hat{X}_i$  while keeping the pre-trained model fixed:

$$\hat{\mathbf{z}} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{(\Theta_j, \mathbf{s}_j)} \mathcal{L}(V(\mathbf{z}_i, \Theta_j; \theta), s_j) + \gamma^2 \|\mathbf{z}\|_2. \quad (13)$$

**Algorithmic Summary.** We summarise the QVF training protocol in Algorithm 1 in the Appendix.

## 5 Experimental Evaluation

We experimentally evaluate our QVF for learning visual field representations, encompassing both 2D images and 3D geometries, while systematically analysing its generalisation in the sense of signal interpolation and the ability to handle missing and occluded regions. We use 1) images from the CIFAR-10 dataset [24] and high-resolution images with rich spectral details [15], and 2) 3D shapes from the ShapeNet [10] dataset. We report widely used metrics averaged over three repetitions.

**Implementation Details.** We empirically evaluate the model on a noiseless high-end simulator: *default.qubit.torch*, provided by PennyLane [6] with an A100 GPU. We employ Adam optimisation [23] with an initial learning rate of  $\eta = 10^{-3}$ , subject to a learning rate scheduler that triggers upon plateauing with a window size of 50 epochs (scaling  $\eta$  by 0.9). The number of epochs is set to  $5k$ , and  $\gamma = 10^{-3}$  in Eq. (12).

**Hardware and Efficiency.** Absence of large-scale, fault-tolerant quantum hardware forces contemporary QML models to rely on exponentially expensive simulators run on classical hardware; see Table 1. For a circuit with  $n$  qubits of depth  $\mathcal{J}$ , the computational complexity on a classical noiseless simulator, without acceleration, is given by  $O(2^{c \cdot n} \mathcal{J})$  where  $c$  is a constant that depends on the specific simulation method employed.

## 5.1 Circuit Trainability

We experimentally show that constraining quantum transformations in  $\hat{S}(\theta)$  to real-valued unitary operations (resulting in bounded Haar randomness) helps with gradient flow. As quantum circuit parameters are inherently periodic within  $[0, 2\pi)$ , we evaluate the gradient flow by uniformly sampling parameters within this range and quantifying its expectation value. Due to the zero-mean nature of the expected loss gradient (see App. A.3), the vanishing gradient phenomenon is governed by the variance decay rate. We, therefore, quantify its variance  $\text{Var}_{\text{grad}}$  as

$$\text{Var}_{\text{grad}} = E_{\theta \sim U([0, 2\pi))} \left[ \text{Var} \left( \left\{ \frac{\partial}{\partial \theta_k} \langle \hat{M} \rangle_{\theta} \right\}_{t=1}^T \right) \right], \quad (14)$$

where “ $\text{Var}(\cdot)$ ” denotes the variance operator and  $T = 500$  is the number of samples to evaluate the expectation.  $\langle \hat{M} \rangle_{\theta}$  is the expectation value of  $\hat{S}(\theta)$ , and  $k$  iterates over ansatz parameters. Fig. 5 reports  $\text{Var}_{\text{grad}}$  for the increasing number of qubits for two ansatz, i.e., of our QVF and QIREN [57] with a strongly-entangled quantum circuit which allows scrambled (i.e., non-restricted) quantum states in the Hilbert space. We observe that our ansatz with bounded Haar randomness maintains a stronger gradient flow, which is crucial for its trainability and efficient representation learning.

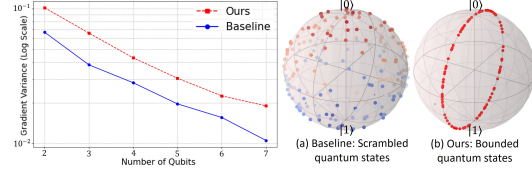


Figure 5: (Left:) comparison of the gradient variance ( $y$ -axis has a log scale); (right:) visualisation of reachable quantum states.

## 5.2 2D (Image) Representation Learning

We evaluate image representation learning with QVF and start with single images. We first compare QVF to a classical model that takes the architecture consistent with QVF’s classical energy inference module; QVF has an overhead of 170 parameters due to its ansatz. This implies that the differences in the learning behaviour and the final representation accuracy are *predominantly* due to the inductive bias of the quantum ansatz, isolating influences from external factors. Fig. 6-(a) visualises reconstructed images during training of QVF; Fig. 6-(b) plots the learning curves (PSNR) for the first thousand training epochs and, thus, highlights the differences in the training progression. Similar observations are made for other trials during the evaluation; QVF significantly accelerates learning high-frequency signals while performing on par with the classical method in the low-frequency regions.

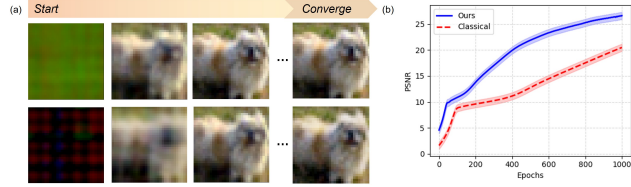


Figure 6: (a) Reconstructed images during training: (top) our QVF, (bottom) classical model; (b) PSNR curves.

We also experiment with the hand-crafted encoding strategy of Rathi et al. [36], which does not result in a recognisable representation upon convergence—an observation consistent with their results. This validates our design and, especially, the necessity of a learnable energy embedding. We then benchmark QVF against QIREN [57], the most closely related QINR approach. While QIREN employs a quantum ansatz sandwiched between classical layers, QVF uses a classical component for data encoding only. We evaluate QVF and QIREN consistently with  $n = 5$  qubits, and evaluate the performance on 50 different images with metrics reported in Table 2. Results demonstrate that QVF with Siren outperforms QIREN by 30% on MSE.

Next, we perform representation learning on image collections via latent variable conditioning, i.e., we configure QVF to learn the 50 images simultaneously. Note that



Figure 7: Visualization of the reconstructed images.

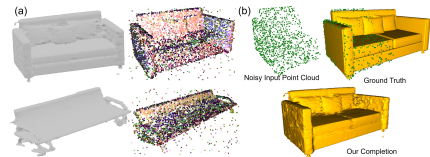


Figure 8: (a): Geometry representation using QVF with latent-space-conditioned SDF inference; (b): Shape completion from partial inputs.

QIREN does not support this experimental setting and, hence, we compare QVF with widely-used classical foundational INR methods, i.e., MLPs with ReLUs, and Siren. The comparisons follow the same evaluation protocol, where the only difference between the baselines and QVF is the presence of the ansatz; the results are summarised in Table 3. Reconstructed images conditioned on different latent codes are visualized in Fig. 7.

We next consider deployment of QVF on future quantum hardware that could reduce the representation fidelity, such as measurement uncertainty. The visual field extracted from the ansatz can differ in its fidelity (accuracy and quality) due to stochastic effects induced by finite sampling on quantum hardware. In App. D, we visualise in Fig. 12 extracted image representations encoded within our pre-trained QVF across a varying number of shots  $N_{\text{shot}}$ , showing the characteristics of the resulting fields with increasing sampling precision.

**Ansatz Configuration.** We next evaluate architectural variations in QVF. We investigate the impact of the key hyperparameters: 1) ansatz width, i.e., number of qubits  $n$ ; 2) circuit depth  $\mathcal{J}$ ; and 3) latent space dimension  $p$  of our classical module for encoding data into quantum states; see Fig. 9 for the results. Scaling up the QVF ansatz, i.e., increasing  $\mathcal{J}$  and width  $n$  (while maintaining other parameters), leads to performance gains in both cases. QVF scales robustly and is not affected by substantial

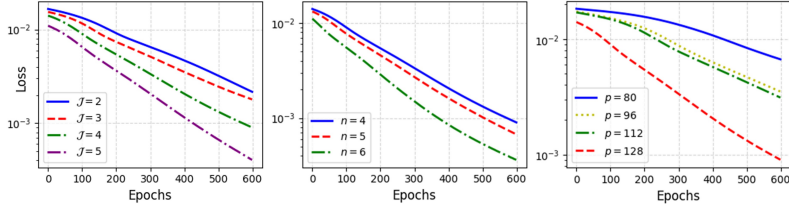


Figure 9: Ablation study with modules influencing the model performance. From left to right: 1) circuit depth  $\mathcal{J}$ ; 2) number of qubits  $n$ ; 3) hidden neuron dimension per layer  $p$ .

trainability problems, at least in our evaluated scenario. The expressivity of the classical module for integrating non-linear data priors and preparing encoded quantum states serves as the architectural cornerstone. With increasing  $p$ , we observe consistent performance surges.

**Parameter Scaling Analysis.** Per design, QVF consists of: 1) the classical module for neural amplitude encoding; and 2) the ansatz. As we leverage a tiny MLP in QVF, its parametrisation scales quadratically, i.e.,  $O(p^2)$  w.r.t. the latent space dimension  $p$ . Meanwhile, our quantum ansatz has parameter scaling of  $O(n\mathcal{J})$  w.r.t. its depth. While the ansatz contributes negligibly to the total parameter count, it improves the overall performance by a large margin (see Fig. 6).

### 5.3 3D (Shape) Representation Learning

We next evaluate geometric representation learning of 3D shapes in the form of SDFs. This setting is investigated for the first time in the context of QINRs; it poses challenges primarily concerning QINR scalability and the complications arising from varying shape topological structures. Similar to images, we perform representation learning on 3D shape collections. We select three shapes from ShapeNet [10] and non-uniformly sample signed distances at  $100k$  spatial points per shape, with higher near-surface sampling density for better surface detail capture. Note that while the scale of this experiment setup can be considered moderate for classical models, it significantly advances the feasible scale of QINR models (which are nevertheless constrained by the simulator’s performance) and provides valuable insights for future advancements. We inherit the experimental setting from the experiments with 2D images and report the representational accuracy upon convergence in Table 3. The baseline setups are, likewise,

| Method              | MSE ( $\times 10^{-3}$ ) $\downarrow$ | PSNR $\uparrow$                   |
|---------------------|---------------------------------------|-----------------------------------|
| Ours(Gaussian)+ReLU | $0.98 \pm 0.09$                       | $30.06 \pm 1.0$                   |
| Ours(Identity)+ReLU | $0.99 \pm 0.09$                       | $30.02 \pm 1.0$                   |
| Ours(Gaussian)+Sin  | $0.55 \pm 0.04$                       | $32.59 \pm 0.2$                   |
| Ours(Identity)+Sin  | <b><math>0.54 \pm 0.04</math></b>     | <b><math>32.67 \pm 0.3</math></b> |
| QIREN [57]          | $0.78 \pm 0.05$                       | $31.03 \pm 0.2$                   |

Table 2: Numerical results for 2D representation learning between the previous QINR method QIREN [57] and our QVF.

| Method              | Images                                |                                  | 3D Shapes                             |
|---------------------|---------------------------------------|----------------------------------|---------------------------------------|
|                     | MSE ( $\times 10^{-3}$ ) $\downarrow$ | PSNR $\uparrow$                  | MAE ( $\times 10^{-3}$ ) $\downarrow$ |
| Ours(Gaussian)+ReLU | <b><math>1.02 \pm 0.11</math></b>     | <b><math>29.8 \pm 1.1</math></b> | <b><math>0.99 \pm 0.07</math></b>     |
| Ours(Identity)+ReLU | $1.03 \pm 0.09$                       | $29.5 \pm 1.2$                   | $1.10 \pm 0.09$                       |
| MLP+ReLU [34]       | $2.17 \pm 0.13$                       | $26.57 \pm 0.51$                 | $1.43 \pm 0.14$                       |
| Ours(Gaussian)+Sin  | <b><math>0.62 \pm 0.05</math></b>     | <b><math>32.2 \pm 0.3</math></b> | <b><math>0.27 \pm 0.05</math></b>     |
| Ours(Identity)+Sin  | $0.72 \pm 0.06$                       | $31.4 \pm 0.4$                   | $0.32 \pm 0.05$                       |
| MLP+Sin [45]        | $1.19 \pm 0.08$                       | $29.2 \pm 0.4$                   | $0.48 \pm 0.06$                       |

Table 3: Numerical results for 2D/3D representation learning for our QVF and classical baselines. The baseline setups are, likewise,

MLPs with different activation functions, including the ReLU, which corresponds to the DeepSDF approach [34]. The final meshes can be extracted from the queried signed distances of QVF using Marching Cubes [27], as visualised in Fig. 8-(a).

#### 5.4 Applications Supported by QVF

QVF supports applications such as visual field interpolation in the latent space, image inpainting and shape completion. Fig. 1-(c) visualises linear latent-space interpolation of 3D shapes encoded in the converged QVF, i.e., a frequent experimental setting in the classical INR literature [34]. QVF also supports image and shape completion by first sampling  $\hat{z}$  and optimising its value using MAP; see details in Sec. 4.3. In the second step, the completion can be performed by leveraging the optimised latent code and inferring missing regions; see Fig. 8-(b) and App. E for the qualitative results.

### 6 Discussion, Future Work and Conclusion

Our QVF is a novel QML framework for implicit representation learning of visual fields. In our experiments on a quantum hardware simulator, we observe that QVF—even with minimal classical components—can achieve high representation fidelity across data modalities such as images and 3D shapes. Furthermore, QVF outperforms the previous QINR approach QIREN (of a similar model scale) and, additionally, is competitive against foundational classical baselines. The ansatz configuration and ablation studies highlight the influence of each QVF module. Our ablative study confirms the sufficient circuit depth resulting in a balance between the ansatz depth and high representational accuracy. Upon our expectations and the theoretical predictions, our QVF is efficient in learning high-frequency signal details (Fig. 6). As the first among QINR methods, QVF supports joint representation learning of image and 3D shape collections, and applications such as image inpainting and 3D shape completion. Finally, we emphasise that this work focuses on the challenges of advancing QINR through fundamental methodological innovations. Hence, we do not aim to challenge classical well-engineered models in the absolute terms. Our implementation can be found on the project page.

**Limitations.** While QVF demonstrates substantial improvements over prior QINR methods in terms of both performance and supported size of visual fields, the current experimental scale, nevertheless, remains constrained due to the quantum hardware simulation overheads. Those, however, affect all existing applied QML works before the advent of fault-tolerant gate-based quantum computers.

**Future Work.** We see various promising avenues for follow-ups and QVF improvements. One direction is to explore the preparation of learnable quantum states following Gibbs distribution with reduced computational complexity (e.g., tensor train decomposition [30]). We also foresee that other problems with open challenges, such as 3D reconstruction and neural rendering from 2D images, could adopt QVF as a representation. We also believe that many tricks and further ideas from the INR literature could be adopted in the QINR context in future (e.g., space partitioning structures and non-rigid generalisations) [38, 49, 46, 47].

**Acknowledgements.** We thank Natacha Kuete Meli, Daniele Lizzio Bosco and Thomas Leimkuehler for helpful comments on the manuscript. The work was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 534951134.

### References

- [1] Amin, M.H., Andriyash, E., Rolfe, J., Kulchitsky, B., Melko, R.: Quantum boltzmann machine. *Physical Review X* **8**(2), 021050 (2018)
- [2] Baek, H., Yun, W.J., Kim, J.: 3d scalable quantum convolutional neural networks for point cloud data processing in classification applications. *arXiv preprint 2210.09728* (2022)
- [3] Ball, C., Cohen, T.D.: Boltzmann distributions on a quantum computer via active cooling. *Nuclear Physics A* **1038**, 122708 (2023)
- [4] Benedetti, M., Lloyd, E., Sack, S., Fiorentini, M.: Parameterized quantum circuits as machine learning models. *Quantum Science and Technology* **4**(4), 043001 (2019)

- [5] Benkner, M.S., Löhner, Z., Golyanik, V., Wunderlich, C., Theobalt, C., Moeller, M.: Q-match: Iterative shape matching via quantum annealing. In: International Conference on Computer Vision (ICCV). pp. 7586–7596 (2021)
- [6] Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., Alam, M.S., Alonso-Linaje, G., AkashNarayanan, B., Asadi, A., et al.: PennyLane: Automatic differentiation of hybrid quantum-classical computations. arXiv preprint 1811.04968 (2018)
- [7] Birdal, T., Golyanik, V., Theobalt, C., Guibas, L.J.: Quantum permutation synchronization. In: Computer Vision and Pattern Recognition (CVPR) (2021)
- [8] Bondarenko, D., Feldmann, P.: Quantum autoencoders to denoise quantum data. Physical review letters **124**(13), 130502 (2020)
- [9] Cerezo, M., Sone, A., Volkoff, T., Cincio, L., Coles, P.J.: Cost function dependent barren plateaus in shallow parametrized quantum circuits. Nature communications **12**(1), 1791 (2021)
- [10] Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet: An Information-Rich 3D Model Repository. arXiv preprint 1512.03012 (2015)
- [11] Chen, D., Yuan, L., Liao, J., Yu, N., Hua, G.: Stylebank: An explicit representation for neural image style transfer. In: Computer Vision and Pattern Recognition (CVPR). pp. 1897–1906 (2017)
- [12] Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: Computer Vision and Pattern Recognition (CVPR) (2021)
- [13] Cong, I., Choi, S., Lukin, M.D.: Quantum convolutional neural networks. Nature Physics **15**(12), 1273–1278 (2019)
- [14] Farina, M., Magri, L., Menapace, W., Ricci, E., Golyanik, V., Arrigoni, F.: Quantum multi-model fitting. In: Computer Vision and Pattern Recognition (CVPR). pp. 13640–13649 (2023)
- [15] Gardner, J.P.: James Webb Space Telescope. <https://webb.nasa.gov/content/multimedia/images.html> (2022)
- [16] Golyanik, V., Theobalt, C.: A quantum computational approach to correspondence problems on point sets. In: Computer Vision and Pattern Recognition (CVPR) (2020)
- [17] Google Quantum AI and Collaborators: Quantum error correction below the surface code threshold. Nature (2024)
- [18] Grant, E., Wossnig, L., Ostaszewski, M., Benedetti, M.: An initialization strategy for addressing barren plateaus in parametrized quantum circuits. Quantum **3**, 214 (2019)
- [19] Haar, A.: Der massbegriff in der theorie der kontinuierlichen gruppen. Annals of Mathematics **34**(1), 147–169 (1933)
- [20] Holmes, Z., Sharma, K., Cerezo, M., Coles, P.J.: Connecting ansatz expressibility to gradient magnitudes and barren plateaus. PRX Quantum **3**(1), 010313 (2022)
- [21] Jospin, L.V., Laga, H., Boussaid, F., Buntine, W., Bennamoun, M.: Hands-on bayesian neural networks—a tutorial for deep learning users. IEEE Computational Intelligence Magazine **17**(2), 29–48 (2022)
- [22] Kerenidis, I., Landman, J., Luongo, A., Prakash, A.: q-means: A quantum algorithm for unsupervised machine learning. Advances in Neural Information Processing Systems (NeurIPS) **32** (2019)
- [23] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint 1412.6980 (2014)
- [24] Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images. University of Toronto (2009)
- [25] Li, Y., Li, S., Sitzmann, V., Agrawal, P., Torralba, A.: 3d neural scene representations for visuomotor control. In: Conference on Robot Learning. pp. 112–123. PMLR (2022)
- [26] Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum principal component analysis. Nature Physics **10**(9), 631–633 (2014)
- [27] Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: Seminal graphics: pioneering efforts that shaped the field, pp. 347–353. ACM (1998)

- [28] McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.: Barren plateaus in quantum neural network training landscapes. *Nature communications* **9**(1), 4812 (2018)
- [29] Meli, N.K., Golyanik, V., Benkner, M.S., Moeller, M.: QuCOOP: A versatile framework for solving composite and binary-parametrised problems on quantum annealers. In: *Computer Vision and Pattern Recognition (CVPR)* (2025)
- [30] Melnikov, A., Termanova, A., Dolgov, S., Neukart, F., Perelshtein, M.: Quantum state preparation using tensor networks. *Quantum Science and Technology* **8**(3), 035027 (2023)
- [31] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021)
- [32] Mitarai, K., Negoro, M., Kitagawa, M., Fujii, K.: Quantum circuit learning. *Physical Review A* **98**(3), 032309 (2018)
- [33] Molaei, A., Aminimehr, A., Tavakoli, A., Kazerouni, A., Azad, B., Azad, R., Merhof, D.: Implicit neural representation in medical imaging: A comparative survey. In: *International Conference on Computer Vision (ICCV)*. pp. 2381–2391 (2023)
- [34] Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: *Computer Vision and Pattern Recognition (CVPR)* (2019)
- [35] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A.: On the spectral bias of neural networks. In: *International conference on machine learning*. pp. 5301–5310. PMLR (2019)
- [36] Rathi, L., Tretschk, E., Theobalt, C., Dabral, R., Golyanik, V.: 3D-QAE: Fully quantum auto-encoding of 3d point clouds. In: *The British Machine Vision Conference (BMVC)* (2023)
- [37] Rebentrost, P., Mohseni, M., Lloyd, S.: Quantum support vector machine for big data classification. *Physical review letters* **113**(13), 130503 (2014)
- [38] Reiser, C., Peng, S., Liao, Y., Geiger, A.: Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In: *International Conference on Computer Vision (ICCV)* (2021)
- [39] Schalkers, M.A., Möller, M.: On the importance of data encoding in quantum boltzmann methods. *Quantum Information Processing* **23**(1), 20 (2024)
- [40] Schleich, P., Skreta, M., Kristensen, L., Vargas-Hernandez, R., Aspuru-Guzik, A.: Quantum deep equilibrium models. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 37, pp. 31940–31967 (2024)
- [41] Schuld, M., Bocharov, A., Svore, K.M., Wiebe, N.: Circuit-centric quantum classifiers. *Physical Review A* **101**(3), 032308 (2020)
- [42] Schuld, M., Sinayskiy, I., Petruccione, F.: An introduction to quantum machine learning. *Contemporary Physics* **56**(2), 172–185 (2015)
- [43] Schuld, M., Sweke, R., Meyer, J.J.: Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A* **103**(3), 032430 (2021)
- [44] Shiba, K., Sakamoto, K., Yamaguchi, K., Malla, D.B., Sogabe, T.: Convolution filter embedded quantum gate autoencoder. *arXiv preprint 1906.01196* (2019)
- [45] Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 33, pp. 7462–7473 (2020)
- [46] Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., Jacobson, A., McGuire, M., Fidler, S.: Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In: *Computer Vision and Pattern Recognition (CVPR)* (2021)
- [47] Tewari, A., Thies, J., Mildenhall, B., Srinivasan, P., Tretschk, E., Yifan, W., Lassner, C., Sitzmann, V., Martin-Brualla, R., Lombardi, S., et al.: Advances in neural rendering. In: *Computer Graphics Forum*. vol. 41, pp. 703–735. Wiley Online Library (2022)
- [48] Thanasilp, S., Wang, S., Nghiem, N.A., Coles, P., Cerezo, M.: Subtleties in the trainability of quantum machine learning models. *Quantum Machine Intelligence* **5**(1), 21 (2023)

- [49] Tretschk, E., Tewari, A., Golyanik, V., Zollhöfer, M., Lassner, C., Theobalt, C.: Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In: International Conference on Computer Vision (ICCV) (2021)
- [50] Tschernezki, V., Laina, I., Larlus, D., Vedaldi, A.: Neural feature fusion fields: 3d distillation of self-supervised 2d image representations. In: International Conference on 3D Vision (3DV). pp. 443–453 (2022)
- [51] Wang, S., Fontana, E., Cerezo, M., Sharma, K., Sone, A., Cincio, L., Coles, P.J.: Noise-induced barren plateaus in variational quantum algorithms. *Nature communications* **12**(1), 6961 (2021)
- [52] Weigold, M., Barzen, J., Leymann, F., Salm, M.: Data encoding patterns for quantum computing. In: Conference on Pattern Languages of Programs (2020)
- [53] Wierichs, D., Izaac, J., Wang, C., Lin, C.Y.Y.: General parameter-shift rules for quantum gradients. *Quantum* **6**, 677 (2022)
- [54] Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., Sridhar, S.: Neural fields in visual computing and beyond. In: Computer Graphics Forum. vol. 41, pp. 641–676. Wiley Online Library (2022)
- [55] Zaech, J.N., Liniger, A., Danelljan, M., Dai, D., Van Gool, L.: Adiabatic quantum computing for multi object tracking. In: Computer Vision and Pattern Recognition (CVPR). pp. 8811–8822 (2022)
- [56] Zhang, K., Liu, L., Hsieh, M.H., Tao, D.: Escaping from the barren plateau via gaussian initializations in deep variational quantum circuits. *Advances in Neural Information Processing Systems* **35**, 18612–18627 (2022)
- [57] Zhao, J., Qiao, W., Zhang, P., Gao, H.: Quantum implicit neural representations. In: International Conference on Machine Learning (ICML) (2024)

# NeurIPS Paper Checklist

## 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The claims in the abstract are detailed and expanded in the main body of the paper.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

## 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: limitations are discussed and included in the main text.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

## 3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: necessary theories are provided in the main text.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.

- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Yes, all necessary details are provided.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

#### 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: all necessary algorithmic details are provided in the main text to re-produce the results. and the source code will be made available when the paper is released.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.

- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

#### 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [\[Yes\]](#)

Justification: training/testing process of the model is detailed in the main text.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

#### 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [\[Yes\]](#)

Justification: related error bars are reported.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

#### 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [\[Yes\]](#)

Justification: this information is provided.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.

- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

#### 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: the code of ethics is strictly complied.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

#### 10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: this work does not have direct social impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

#### 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: we did not find high risk for misuse with our work.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: we credited external codes

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

## 13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: we do not provide new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

## 14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: it is not provided.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

## 15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: no human subjects are involved.888

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

#### 16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: No LLMs were involved except for writing.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

# Supplementary Material

This appendix supplements the main paper, starting in Sec. A with a detailed background on gate-based quantum computing. It covers both quantum physics foundations and applications of quantum computing in machine learning and INR. It also connects quantum circuit measurements and Bayesian inference. Next, we outline the full algorithmic (training) protocol; visualise the quantum ansatz architecture used in the experiments; and provide more implementation details in Sec. B. On the experimental side, we further analyse QVF performance with noisy circuits in Sec. C, followed by the image representation quality in dependence on the number of measurement repetitions (shots) in Sec. D. Applications supported by QVF, such as image inpainting and shape completion, are discussed in Secs. E and F, while additional visualisations of 3D shapes are shown in Sec. G. Sec. H then discusses the development status of currently available real quantum hardware.

## A Background

### A.1 Preliminaries on 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  $|1\rangle$ .

*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:

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \quad (15)$$

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:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, |\psi\rangle = \alpha |0\rangle + \beta |1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (16)$$

*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  $|1\rangle$ :

$$\hat{O} |0\rangle = \kappa |0\rangle \text{ and } \hat{O} |1\rangle = \delta |1\rangle, \quad (17)$$

where  $\kappa$  and  $\delta$  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.,  $\kappa$  or  $\delta$ .

*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:

$$|\psi\rangle_2 = a |00\rangle + b |01\rangle + c |10\rangle + d |11\rangle, \quad (18)$$

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.,

$$|\psi\rangle_2 \neq |\psi\rangle_{a1} \otimes |\psi\rangle_{a2}. \quad (19)$$

**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  $\theta, \tau$  and  $\gamma$ :

$$\hat{R}_x(\theta) = \begin{bmatrix} \cos(\frac{\theta}{2}) & -i \sin(\frac{\theta}{2}) \\ -i \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{bmatrix}, \quad \hat{R}_y(\tau) = \begin{bmatrix} \cos(\frac{\tau}{2}) & -\sin(\frac{\tau}{2}) \\ \sin(\frac{\tau}{2}) & \cos(\frac{\tau}{2}) \end{bmatrix}, \quad \hat{R}_z(\gamma) = \begin{bmatrix} e^{-i\frac{\gamma}{2}} & 0 \\ 0 & e^{i\frac{\gamma}{2}} \end{bmatrix}. \quad (20)$$

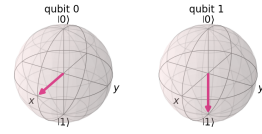


Figure 10: Bloch sphere visualisation of qubit states. Qubit 0:  $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ , qubit 1:  $|\psi\rangle = |1\rangle$ .

The Pauli operators  $\hat{X}, \hat{Y}, \hat{Z}$  represent specific instances of above rotation operators, i.e. rotations by  $\pi$  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:

$$\hat{X} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \hat{Y} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \hat{Z} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (21)$$

**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:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H}(t) |\psi(t)\rangle, \quad (22)$$

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:

$$|\psi(t)\rangle = \hat{T} e^{-\frac{i}{\hbar} \int_0^t \hat{H}(t) dt} |\psi(0)\rangle, \quad (23)$$

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:

$$|\psi(t)\rangle = \hat{U}(\hat{H}, t) |\psi(0)\rangle, \text{ with} \quad (24)$$

$$\hat{U}(\hat{H}, t) = e^{-\frac{it}{\hbar} \hat{H}}. \quad (25)$$

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:

$$\hat{U}(\hat{H}, t) = e^{-\frac{it}{\hbar} \hat{H}} = e^{-\frac{i\eta}{2} \hat{\sigma}} = \hat{R}_{\sigma}(\eta). \quad (26)$$

## A.2 Review: Quantum Machine Learning

The potential of quantum computing to enhance machine learning algorithms leads to the emergence of quantum machine learning (QML) [42], a discipline employing quantum mechanical phenomena to tackle classically intractable learning problems through enhanced computational paradigms [42]. 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 [4, 43]. Several standardized QML algorithms have been explored, including quantum principal component analysis [26], quantum support vector machines [37], quantum Boltzmann machines [1], and quantum k-means clustering [22].

**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  $\epsilon$ ) 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  $\hat{U}(\theta) = \mathcal{T} \left( \prod_{i=1}^t \hat{U}_i(\theta_i) \right)$ , where  $\mathcal{T}$  denotes the time-ordering operator governing gate sequence implementation. This induces a Hilbert space transformation  $\hat{U}(\theta) : \mathcal{H} \rightarrow \mathcal{H}$  that maps input states to processed output states through the operation  $|\phi(x, \theta)\rangle = \hat{U}(\theta) |\psi(x)\rangle$ .

**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 [53]. To minimize a measurement-dependent cost

function  $\mathcal{L}(\theta)$ , the exact gradients  $\nabla \mathcal{L}(\theta)$  can be evaluated through quantum circuit evaluations at shifted parameters  $\theta \pm \frac{\pi}{2} \mathbf{e}_i$  for basis vectors  $\mathbf{e}_i$ , expressed as:

$$\partial_{\theta_i} \mathcal{L} = \frac{1}{2} \left[ \mathcal{L} \left( \theta_i + \frac{\pi}{2} \right) - \mathcal{L} \left( \theta_i - \frac{\pi}{2} \right) \right]. \quad (27)$$

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.3 Review: Barren Plateaus

Training a variational ansatz  $\hat{S}(\theta)$  is fundamentally constrained by the *barren plateau* phenomenon, where random parameter initialisation induces exponential vanishing of cost function gradients across Hilbert space. As formally demonstrated by McClean et al. [28] through concentration of measure analysis:

*“...for a wide class of reasonable parametrised 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 phenomenon is also known as *barren plateau*, which can be expressed mathematically for a system with  $n$  qubits as follows:

$$\mathbb{E}_w[\partial_w L(w)] = 0, \quad \text{Var}_w[\partial_w L(w)] \in O\left(\frac{1}{\nu^n}\right), \quad \nu > 1, \quad (28)$$

where  $\nu$  characterises the circuit’s entangling capacity. The variance bound’s scaling establishes that gradient estimators require  $O(\nu^n)$  measurement samples to maintain constant precision, resulting in an exponential resource overhead that renders practical optimisation infeasible for  $n \gg 1$ . This poses challenges, particularly for gradient-based learning. Identified factors contributing to barren plateaus include observable locality [9, 48], specific noise models [51], and an ansatz close to a 2-design, i.e., matching Haar random unitaries up to the second moment [28, 20]. Those highlight the importance of selecting appropriate initialisation protocols, quantum ansatz designs and observables.

### A.4 Connection of PQC to Bayesian Inference

As quantum circuits are inherently probabilistic models, they share conceptual parallels with Bayesian inference. In Bayesian neural networks (BNNs), probabilistic outputs emerge from parameters governed by prior distributions, with training focused on maximising 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 PQC 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 [21]. This distinction underscores that the Bayesian interpretation of PQCs arises from their measurement statistics rather than an intrinsic probabilistic parameter space.

## B Algorithmic Protocol and Ansatz

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

### B.1 Detailed Variational Ansatz Visualisation

We also visualise our QVF ansatz with bounded Haar randomness. Fig. 11 compares the reachable Hilbert states between ours and the ansatz in QIREN that does not restrict the set of possible operations to real-valued unitaries [41, 57]. We highlight the circuit structures via their parametrised single-qubit rotations and inter-qubit entanglement patterns. Traversable quantum states are visualised on the Bloch sphere by sampling the ansatz.

### B.2 Additional Implementation Details

We next provide additional implementation details on our experimental setup.

**Parameterisation of QVF vs. QIREN.** QVF is designed to be a compact QINR model. Compared to QIREN, QVF does not need classical post-processing while maintaining high representational accuracy. The number of parameters in our experiments is  $0.52 \cdot 10^5$  (corresponding to  $p = 128$ ). vs.  $0.74 \cdot 10^5$  for QVF and QIREN, respectively. The ansatz of QVF is configured with  $\mathcal{J}=5$  and  $n=5$ , and for QIREN, we use the default depth.

---

**Algorithm 1** QVF Training Protocol

---

- 1: **Input:** Training dataset  $X = \{(\Theta_i, s_i)\}_{i=1}^W$ ; number of qubits  $n$ ; epochs  $N_{\text{epoch}}$ ; measurement shots  $N_{\text{shot}}$ ; parameters  $\theta = \{\theta_q, \theta_c\}$ ; inverse temperature  $\beta$ .
- 2: **for** epoch = 1 **to**  $N_{\text{epoch}}$  **do**
- 3:   **Classical Inference (Sec. 4.1):**
- 4:     Compute energy spectrum  $\mathbf{E}(\Theta_i; \theta_c)$ .
- 5:     Evaluate Gibbs distribution:

$$P_i = \frac{e^{-\beta E_i}}{Z}, \text{ where } Z = e^{-\beta E_j}.$$

- 6:   **Quantum State Preparation:**
- 7:     Initialise  $\hat{\rho}_0 = \sum_{i=1}^{2^n} P_i |i\rangle \langle i|$ .
- 8:   **Quantum Evolution (Sec. 4.2):**
- 9:     Apply ansatz  $\hat{S}(\theta_q) = \prod_{\ell=1}^J e^{-i\theta_{q,\ell} \hat{H}_\ell}$  to obtain:

$$\hat{\rho}(\theta_q) = \hat{S}(\theta_q) \hat{\rho}_0 \hat{S}^\dagger(\theta_q).$$

- 10:   **Measurement and Observables:**
- 11:     Estimate  $\langle \hat{M}_k \rangle = \text{Tr}[\hat{\rho}(\theta_q) \hat{M}_k]$  for  $k = 1, \dots, K$ .
- 12:   **Gradient Computation of the Loss  $\mathcal{L}$  (Sec. 4.3):**
- 13:     Quantum:  $\partial \langle \hat{M}_k \rangle / \partial \theta_q$  via parameter-shift rule.
- 14:     Classical:  $\nabla_{\theta_c} \mathcal{L}$  via automatic differentiation through  $\mathbf{E}(\Theta_i; \theta_c)$ .
- 15:   **Parameter Update:**
- 16:     Adam optimiser step with learning rate  $\eta$ :

$$\theta \leftarrow \theta - \eta (\nabla_{\theta_c} \mathcal{L}, \nabla_{\theta_q} \mathcal{L}).$$

17: **end for**

18: **Output:** Optimised parameters  $\theta_q^*, \theta_c^*$ .

---

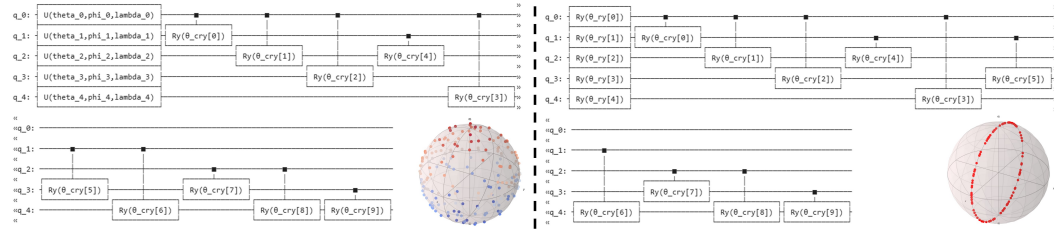


Figure 11: Visualisation of ansatz designs and their exemplary induced traversable quantum states within Hilbert space: strongly entangled ansatz (left) and QVF (right). Traversable states of different circuit ansatzes are visualised on the bottom right for both ansatzes.

**Simulation Stability in Preparing Quantum States following Gibbs distribution.** When simulating the preparation of such quantum states, the partition function evaluation involves exponentiations of large and small Hamiltonian eigenvalues; see Eqs. (2) and (3), which could cause numerical instabilities. We leverage the log-sum-exp trick, a well-established numerical stabilisation technique for this problem.

## C Image Representation with Noisy Circuits

Evaluation with quantum circuit noise can provide valuable insights for the practical deployment of QVF on near-term quantum hardware. Hence, we investigate the influence of quantum gate infidelities on the performance of QVF, i.e., a dominant source of errors and noise in the quantum operations. Gate operation infidelity arises

| Method                 | no noise        | $\sigma = 0.01$ | $\sigma = 0.05$ | $\sigma = 0.1$  |
|------------------------|-----------------|-----------------|-----------------|-----------------|
| Ours (Gaussian) + ReLU | $30.06 \pm 0.1$ | $30.1 \pm 0.1$  | $28.42 \pm 0.1$ | $25.78 \pm 0.2$ |
| Ours (Identity) + ReLU | $30.02 \pm 0.2$ | $29.6 \pm 0.1$  | $27.98 \pm 0.2$ | $25.96 \pm 0.2$ |
| Ours (Gaussian) + Sin  | $32.59 \pm 0.2$ | $32.4 \pm 0.2$  | $30.66 \pm 0.1$ | $27.94 \pm 0.2$ |
| Ours (Identity) + Sin  | $32.67 \pm 0.3$ | $32.8 \pm 0.2$  | $30.34 \pm 0.2$ | $28.12 \pm 0.2$ |

Table 4: QVF performance with noisy circuits.  $\sigma$  is the perturbation ratio modelling quantum circuit infidelities.

from intrinsic control imperfections in quantum hardware, resulting in stochastic deviations of the performed gate operations from their expected behaviour. These imperfections constrain gate fidelity to finite precision, which can be effectively modelled as zero-mean perturbations to the gate parameters within a bounded range. To simulate the impact of such noise, we introduce zero-mean Gaussian perturbations with varying standard deviations: Higher values correspond to the noise levels typical for current near-term quantum devices, while lower values reflect anticipated improvements in future hardware. The experiments are performed on selected 2D images using the same quantum hardware simulator of PennyLane [6] as in the main matter (Sec. 5.2). We report the results with different levels of quantum gate perturbation ratio  $\sigma$  in Table 4 to quantify the degradation in performance under various noise regimes. As expected, increasing  $\sigma$  leads to a decrease in 2D image PSNR. Even with  $\sigma = 0.1$ , we achieve a PSNR of  $\sim 25$ dB or higher (cf. Fig. 12).

## D Image Representation with a Different Number of Samples (Shots)

QINR of images under finite sampling is fundamentally governed by the statistical uncertainty inherent to quantum measurement. The image quality depends on the number of shots, i.e., QINR query repetitions. We visualise the influence of the number of shots  $N_{\text{shot}}$  (in total per image) in our ansatz in Fig. 12 through the progressive reduction of shot noise artefacts for an increasing number of shots from 100 to  $10^4$ . With low measurement shots, zero-mean sampling noise dominates the representation. As number of shots  $N_{\text{shot}}$  increase and approaches  $10^3$ , noise suppression becomes significant as expected according to CLT as noise variance follows  $\sigma^2 \propto 1/N_{\text{shots}}$ , allowing the representation to better approximate the ground truth. The observed noise patterns across the different shot numbers are characteristic of the proposed ansatz of QVF and will serve as a reference for future research.

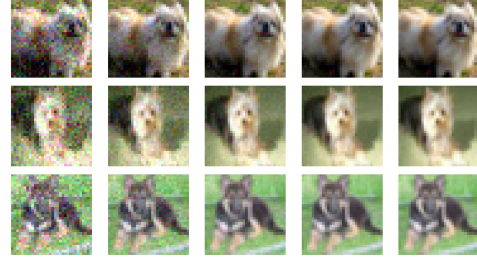


Figure 12: Qualitative images retrieved from a pre-trained QVF under the different number of circuit shots. From left to right, shot counts are 100, 500,  $10^3$  and  $10^4$ , respectively. The rightmost images represent the ground truth.

## E Application: Image Inpainting

QIREN [57] and 3D-QAE [36] are constrained by their reliance on fixed latent representations or rigid interpolation mechanisms, thereby being incapable of reconstructing 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, enabling applications such as image inpainting. Given images with occluded or corrupted pixels, the circuit identifies a vector in the latent space that minimises the discrepancy between the predicted multi-dimensional properties learned by the quantum circuit and the observed noisy values. The optimised latent vector conditioning the quantum circuit enables recovery of missing field properties. Empirically, we masked out half of the image pixels and reconstructed the complete images via the protocol. Representative results of image inpainting with QVF pre-trained on 50 images are visualised in Fig. 13, 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 these tasks.

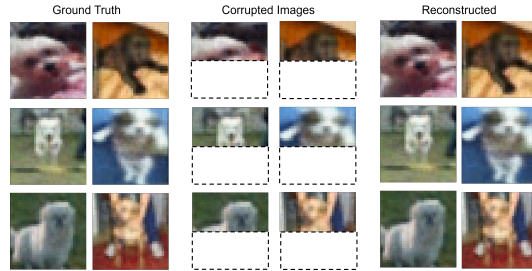


Figure 13: Image inpainting results with QVF.

## F Application: Shape Completion from Partial and Noisy Depth Maps

Similarly to image inpainting, QVF can be used for tasks such as 3D geometry completion given noisy depth maps. We adopt a similar setting as for images by cropping half of the samples along the depth dimension. We then study the effects of zero-mean Gaussian noise applied to the depth maps across different perturbation ratios  $\alpha$ ; see Fig. 8-(b). Shape completion performance is quantified across incremental perturbation ratios, parameterised as  $\alpha \in \{0, 0.005, 0.01, 0.02, 0.03\}$ , where  $\alpha = 0$  corresponds to an idealised noise-free scenario. This qualitative analysis reveals a monotonic decline in reconstruction fidelity with increasing noise, as evidenced by progressive geometric distortions and surface irregularities in Fig. 12.

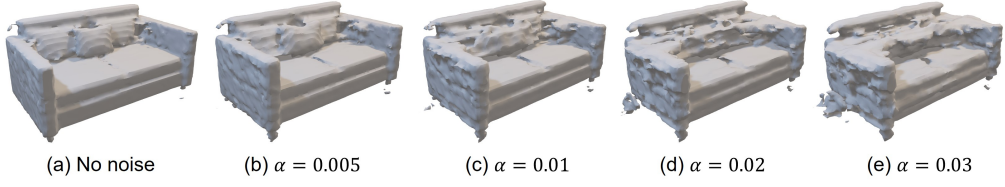


Figure 12: Shape completion from partial and noisy input depth maps using QVF;  $\alpha$  is the noise ratio.

## G 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.

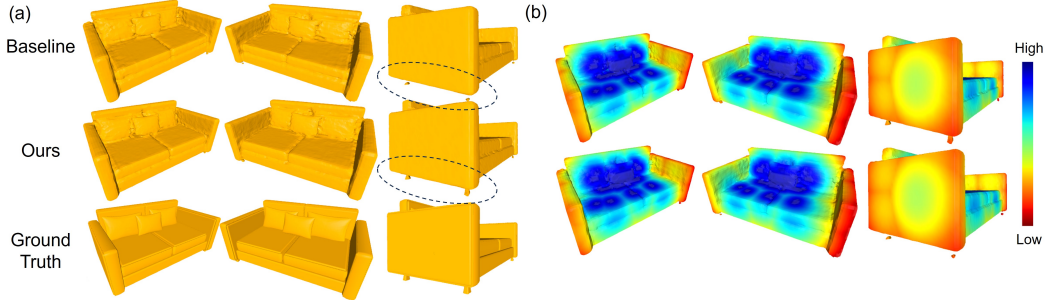


Figure 13: (a) Comparison of geometric representations using QVF and the classical model; ground truth is presented at the bottom; (b) representation fidelity visualized via color-encoded Hausdorff distance map: colors represent the distance to the ground truth. The rendered image employs a color gradient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

## H Existing Gate-based Quantum Hardware

Same as the prior work [57], we evaluate the proposed QVF on a simulator [6] due to the immaturity of real quantum hardware for high-level and practical visual computing tasks. Existing gate-based quantum platforms—including superconducting circuits, trapped ions, neutral atoms, photonic systems, and quantum dots—are in varying stages of development, with none yet achieving the maturity required for large-scale, fault-tolerant computation. Limiting factors include noise susceptibility, restricted execution time due to quantum decoherence, and the necessity for error correction. Nevertheless, as classical machine learning systems continue to expand in scale, driving unprecedented computational and energy demands, rapid advancements in quantum computing techniques and hardware [17] are anticipated to address these barriers in the foreseeable future, underscoring the need to proactively explore applications executable on emerging quantum computers such as QVF.