How Much Does the Quantum Part Do? Auditing Quantum Implicit Neural Representations

Anonymous authors
Paper under double-blind review

Abstract

Quantum implicit neural representations place a parametrised circuit inside a coordinate network and report gains over classical baselines. We rebuild two of them under a common evaluation harness at matched parameter budgets, recover the reported performance level for both, and then ask what produces it. For Quantum Visual Fields the answer is almost entirely classical: changing one activation in the classical embedder is worth 5.52 dB, while training the 80-parameter ansatz, 0.17% of the model, is worth -0.05 dB at width 128 and $+1.48$ dB at width 256, the largest effect we observe and the only multi-seed configuration whose paired 95% confidence interval excludes zero, at 40–$75\times$ the training cost. A proposition characterises the readout: the circuit's output is a quadratic form in the embedder's own output; each channel's form lies in a 136-dimensional symmetric-matrix space, while the jointly reachable family has intrinsic dimension at most 80. A budget-matched classical quadratic readout holding the same 80 trainable parameters matches the circuit at width 128 and beats it by 2.85 dB at width 256, the very configuration where training the ansatz helps most, while training three orders of magnitude faster. The same bottleneck is consistent with an empirical finding, that the architecture is flat in classical width while its one effective scaling direction costs 2^n. At matched parameter budgets SIREN leads in fitting quality, while QIREN requires three to four orders of magnitude more compute. We also document that the released artifact does not run as shipped and omits the variant producing its published numbers.

1 Introduction

Implicit neural representations (INRs) encode a signal as a coordinate-to-value network (Sitzmann et al., 2020; Tancik et al., 2020; Mildenhall et al., 2020). The classical side of the field has moved quickly: multiresolution hash encodings (Müller et al., 2022), band-limited networks (Lindell et al., 2022), wavelet (Saragadam et al., 2023), Gaussian (Ramasinghe & Lucey, 2022) and variable-periodic (Liu et al., 2024) activations, and a unifying account of the sinusoidal family (Morsali et al., 2026); see Essakine et al. (2025) for a survey. A recent line of work replaces part of that network with a parametrised quantum circuit, producing *quantum* implicit neural representations (QINRs) (Zhao et al., 2024; Wang et al., 2025). The most advanced of these, Quantum Visual Fields (QVF) (Wang et al., 2025), learns an energy-based amplitude encoding, evolves it through a fully entangled ansatz, and reads out visual fields without classical post-processing. It reports improvements over a previous QINR and over classical INR baselines on 32×32 image collections and on ShapeNet signed distance fields. We audit the image half; our shape runs did not reproduce the published values and we exclude them (§C).

Claims of this form are difficult to assess from published tables alone, because the comparison depends on three quantities that are rarely reported together: how many parameters each model is given, how much tuning each baseline received, and how much of the model is actually quantum. This paper supplies those quantities for QVF.

Our method is deliberately narrow. We take the authors' released code, fix what is needed to make it run, and place QVF and its comparators in a harness that trains every arm on the same data, with

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the same latent conditioning, optimiser and epoch budget, at matched parameter counts. We later evaluate QIREN in the same harness under its own prescribed training settings. Within that harness we ask three questions. First, what do the quantum parameters contribute? We freeze exactly the ansatz parameters and leave everything else identical, so the difference between the two arms is the contribution of training them, with no confound from changing architecture. Second, how do properly configured classical baselines compare? We sweep the one hyperparameter that dominates periodic-activation INRs, the first-layer frequency scale ω_0, and match parameter counts. Third, how does each family scale? We enlarge the classical width and the quantum ansatz, in depth J and in qubit count n, and compare the trends.

Contributions. (i) A reproducibility audit of the released QVF artifact, including two independent blocking defects that prevent a single forward pass, a training protocol described in the paper but absent from the code, and 19 configuration keys the code never reads (§3). (ii) A matched-protocol, matched-parameter re-evaluation of the image modality, with a frozen-ansatz ablation that isolates the quantum contribution (§6). (iii) The same harness applied to a second QINR, QIREN, run under its own prescribed optimiser settings, whose published number we also recover (§7), and a scaling comparison showing that classical width scales cleanly while neither quantum ansatz does (§B). (iv) Two propositions characterising the ansatz — that identity-block initialisation makes it exactly the identity, and that its output is a quadratic form whose matrix lies, per channel, in a 136-dimensional symmetric-matrix space while the jointly reachable family has intrinsic dimension at most 80 — and the classical surrogates the second one implies, which match or beat the quantum model at three orders of magnitude less compute (§4). (v) An explicit statement of what we could not establish: our shape-modality runs matched neither the published values nor the published ordering of the two embedder activations, and we exclude that half rather than report a reconstruction we cannot vouch for (§C).

2 Background

A QINR maps a coordinate $\boldsymbol{x}$, optionally with a per-signal latent code $\boldsymbol{z}$, to a signal value through a hybrid pipeline. In QVF a classical MLP f_θ consumes a positional encoding (Tancik et al., 2020; Mildenhall et al., 2020) of $\boldsymbol{x}$ concatenated with $\boldsymbol{z}$ and emits an energy vector; a Boltzmann transform turns it into a probability vector whose square root is loaded as the amplitudes of an n-qubit state, a parametrised circuit $\hat{S}(\boldsymbol{\theta}_q)$ evolves that state, and a fixed parity map combines the Pauli-Z expectations into the output channels. Training is end-to-end through the simulator.

The expressivity of such models is governed by the frequency spectrum the encoding makes available (Schuld et al., 2021), the same mechanism that Fourier-feature and periodic-activation INRs exploit (Tancik et al., 2020; Sitzmann et al., 2020). Whether it survives contact with a classical competitor is a live question: circuits shallow enough to avoid barren plateaus (McClean et al., 2018) are often classically simulable (Cerezo et al., 2025; Gil-Fuster et al., 2025), classical surrogates can be extracted from trained quantum learners (Schreiber et al., 2023; Nair & Ferrie, 2025), and the diagnostics used to establish a quantum contribution are themselves contested (Aghaei Saem et al., 2026; Gil-Fuster et al., 2024; Kesiku & Garcia-Zapirain, 2026), with attribution audits now appearing in other domains (Gillani et al., 2026). The relevant comparison is therefore a correctly configured classical INR (Essakine et al., 2025), and configuration matters: SIREN's advantage rests on a specific initialisation coupled to a first-layer frequency scale ω_0, without which periodic activations perform poorly.

3 Reproducibility Audit

We built the environment pinned by the authors' `environment.yaml` (PennyLane 0.34.0 (Bergholm et al., 2018), PyTorch 2.1.2, torchvision 0.16.2) and attempted to run the released scripts unmodified.

The artifact does not execute as shipped. Two independent defects block execution. In the image pipeline, the dataset truncates its image tensor to `num_samples` but then reshapes the *untruncated* tensor, so construction fails for any configuration in which the requested count differs from the

number of available images — including the shipped one. In the same pipeline, the quantum layer allocates its rotation parameters as a rank-2 tensor but indexes them with three subscripts inside the circuit, so *both* initialisation branches raise an `IndexError` before any gradient is computed. The shape pipeline reads a configuration key (`parity_channels`) that the shipped configuration does not define, and requests five training shapes where the repository ships three.

The intended semantics are recoverable. The shape pipeline contains the same circuit written correctly (`weights[:n_qubits, i]`). We adopt that indexing verbatim in the image pipeline; the identity-block initialisation then behaves as designed, with the forward and adjoint rotations cancelling at initialisation. This gives us confidence that our patch restores the authors' intent rather than substituting our own model.

The released code does not implement the described protocol. The paper's text specifies a learning-rate scheduler triggered on plateau with a window of 50 epochs and a factor of 0.9; the released configuration provides a factor of 0.9 and a patience of 15. The two do not agree with each other, and neither script constructs a scheduler at all. We adopt the configuration's values (factor 0.9, patience 15) for every arm, and note the discrepancy rather than choosing between the two on the authors' behalf. More broadly, the image configuration exposes eight keys the code never reads (including `gaussian_initialization`, both weight-decay terms, and the separate quantum learning rate) and the shape configuration exposes eleven. No evaluation code is included: neither script computes PSNR or MAE, defines a held-out split, or implements any of the baselines named in the paper's tables.

The released code does not implement the reported best variant. The original paper's tables distinguish four QVF variants by initialisation (Gaussian or identity block) and by the activation used inside the classical embedder (ReLU or $\sin$). Both released pipelines hard-code ReLU; the only occurrences of $\sin$ are in the positional encoding. With the shipped configuration (`use_identity_block: True`) the released model is therefore the *Identity+ReLU* variant, which the paper reports at 29.5 dB and 1.10×10^{-3}. The *Gaussian+Sin* variant that produces the paper's best figures (32.2 dB, 0.27×10^{-3}) cannot be obtained from the release without modifying the model. We therefore take Identity+ReLU as the reproduction target, and separately construct the $+\sin$ variant ourselves.

The quantum fraction is small. With the paper's stated width ($p = 128$), the image model holds $46{,}736$ classical parameters against 80 quantum ones (0.17%). In the shape pipeline, which we do not evaluate but which is part of the same release, the split is $211{,}472$ against 20 (0.009%). The original paper attributes observed differences "predominantly" to the inductive bias of the ansatz.

4 What the Circuit Can Do

Before measuring the ansatz we characterise it. Write e_θ for the classical embedder, which emits a Boltzmann distribution $p(x) \in \Delta^{d-1}$ over $d = 2^n$ outcomes and loads its square root as amplitudes, $\psi(x) = \sqrt{p(x)}$; write $U(\boldsymbol{\theta}_q)$ for the circuit, Z_i for the Pauli-Z observable on wire i, and $P \in \mathbb{R}^{C \times n}$ for the fixed matrix by which the released parity routine turns the n measurements into the C output channels. Derivations and the numerical checks are in §E.

Proposition 1 (The frozen arm is classical)**.** *Under identity-block initialisation, $U(\boldsymbol{\theta}_q) = \mathbb{I}$ exactly. The frozen-ansatz model is therefore a function of $p(x)$ alone.*

This is what makes the ablation of §6 valid: the frozen arm is not a quantum model with fixed parameters but a purely classical one, so any difference between the two arms is attributable to the ansatz and to nothing else.

Proposition 2 (The circuit selects a quadratic form)**.** *$f_c(x) = \psi(x)^\top M_c\, \psi(x)$, where $M_c = \sum_{i=1}^{n} P_{ci}\, U^\top Z_i U$ is real symmetric and independent of x. The map $\boldsymbol{\theta}_q \mapsto (M_1, \ldots, M_C)$ is real-analytic, so its image has dimension at most $\dim \boldsymbol{\theta}_q$ inside $(\mathrm{Sym}_d)^C$.*

For the configuration under audit $n = 4$, $d = 16$, $C = 3$ and $\dim \boldsymbol{\theta}_q = 80$. The released parity routine pads the $n = 4$ measurements to a multiple of C and divides by $\lfloor n/C \rfloor = 1$, so it sums

rather than averages, giving

$$P = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad R = z_0 + z_3, \quad G = z_1, \quad B = z_2.$$

Each M_c then lives in a 136-dimensional symmetric-matrix space, the triple lives in a 408-dimensional one, and the jointly reachable family has intrinsic dimension at most 80. We verify the whole map numerically, including that the RGB output and its gradient agree with $\psi^\top M_c \psi$ and $2M_c\psi$ to 10^{-7} (§E). The proposition constrains the readout, not the embedder: $\psi_\theta(x)$ may still grow more expressive as the classical network widens, so this is not by itself a ceiling on the model. What it does give us is a construction, and that is how we use it.

A surrogate the proposition hands us. Proposition 2 specifies a classical *relaxation* of the circuit: keep the embedder, discard the circuit, and learn $M_1, \ldots, M_C$ directly and without constraint. This is a bilinear layer whose hypothesis set strictly contains the circuit's, a superset rather than the same class. We store a full $d \times d$ matrix per channel for convenience and symmetrise it in the forward pass, so the layer carries $nd^2 = 1{,}024$ stored parameters, of which $n\binom{d+1}{2} = 544$ survive symmetrisation and, since $\operatorname{rank} P = 3$, only $3\binom{d+1}{2} = 408$ affect the output: the readout sees M_0 and M_3 solely through their sum. The stored count is what Table 5 reports, which is why the surrogate shows 1,024 parameters more than the frozen arm. Because $\max_{M \in \mathcal{M}_{\text{unrestricted}}} F(M) \geq \max_{\boldsymbol{\theta}_q} F(M(\boldsymbol{\theta}_q))$, the relaxation's *optimum* bounds the circuit's; what we report is not that optimum but the value stochastic gradient descent reaches inside the larger set, so we describe it as an optimisation-accessible superset benchmark rather than an upper bound. A gap in its favour shows that the constraint is not being paid for, not that the circuit was trained badly. Table 5 reports it: with the $\sin$ embedder the relaxation is 0.78 dB ahead over three seeds each (paired 95% CI $[+0.35, +1.20]$) and faster by roughly three orders of magnitude; with the ReLU embedder the two are level.

A reader may object that a superset should win because it has more capacity, so we also ran a surrogate whose learnable readout budget equals the ansatz's exactly. Fixing a random 20-dimensional subspace of Sym_d, shared across the $n = 4$ measurement outputs, and learning only the coefficients gives $4 \times 20 = 80$ trainable readout parameters and a total of 46,816, the same figures as the circuit. By the same rank argument these 80 act through 60 effective directions, so the surrogate is if anything the more constrained of the two. It reaches 33.85 ± 0.56 dB against the circuit's 33.06 ± 0.18, a paired $+0.79$ dB whose interval $[-0.23, +1.81]$ contains zero at three seeds, and it is indistinguishable from the unrestricted relaxation ($+0.02$, $[-0.99, +1.02]$). The extra capacity of the unrestricted relaxation is therefore not necessary to match the circuit: an 80-parameter classical readout drawn from an arbitrary subspace already reaches comparable performance. Only three seeds are available for the paired comparisons because the quantum arm costs 15 hours per seed.

Width 128 is, however, the width at which the ansatz does least, so we repeated both surrogates at width 256, where its measured contribution is largest. The gap widens rather than closes. The budget-matched readout reaches 36.79 ± 0.46 dB against the circuit's 34.63 ± 1.01, a paired $+2.85$ dB with interval $[+1.46, +4.24]$, and the unrestricted relaxation reaches 36.47 ± 0.55. So at the one configuration where training the ansatz buys a measurable 1.48 dB, a classical readout holding the same 80 parameters buys about twice that, in minutes rather than in fifteen hours. The budget-matched readout is also slightly *ahead* of the unrestricted one here ($+0.32$, $[+0.08, +0.56]$), so the additional 348 effective directions are not merely unnecessary but mildly harmful.

Since the subspace is drawn at random, the result could depend on the particular draw. Holding the training seed fixed and redrawing the subspace eight times gives 33.80 ± 0.37 dB over a range of 33.09 to 34.39: seven of the eight draws exceed the circuit's best of three seeds, 33.23 dB, and the weakest draw matches its mean. The three runs in Table 5 vary the subspace and the training seed together, so the ± 0.56 reported there is the wider and more conservative figure.

On fitted coordinates, then, the constraint is not paying for itself: the circuit selects a lower-dimensional subset of a function class that is already classical and already cheaper to search.

Proposition 3 (QIREN's budget is capped by simulation)**.** *QIREN allocates one qubit per hidden unit, so a layer of width h carries a state of 2^h amplitudes: exact state-vector simulation requires $\Theta(2^h)$ memory and $\mathrm{poly}(h)\, 2^h$ time per sample, while the parameter count is $\Theta(h^2)$. Attaining a*

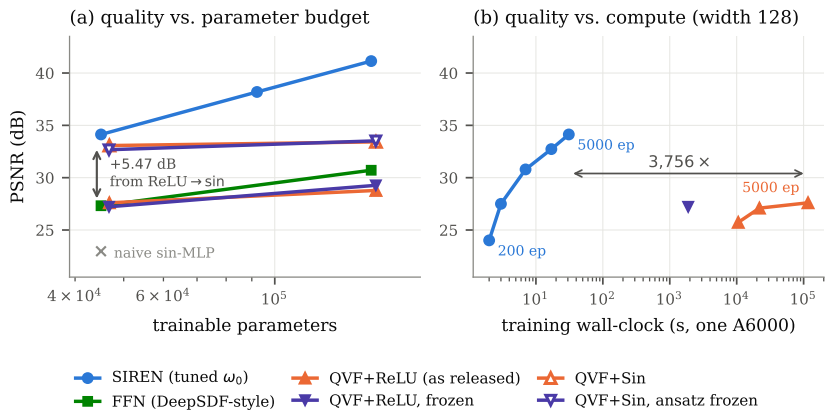

Figure 1: Image collections (50 CIFAR-10 images, latent-conditioned, one protocol). **(a)** Filled markers are the released ReLU variant, open markers the $+\sin$ variant the paper reports as its best. The arrow is the gain from that one classical activation change; the quantum arms are otherwise flat in width while SIREN climbs; the seeded frozen-arm estimate of the same change is 5.52 dB (§6). **(b)** Against wall-clock: SIREN completes its entire $200 \to 5000$ epoch trajectory before QVF's first measured point. Single seed-0 runs, as in Table 1.

budget of P parameters therefore costs $\Theta(2^{\sqrt{P}})$ memory and $\mathrm{poly}(\sqrt{P})\, 2^{\Theta(\sqrt{P})}$ time, against $\Theta(P)$ for a classical network of the same budget.

The constant is not favourable. Measuring peak memory for one training step on 1024 coordinates, a 675-parameter QIREN ($h = 8$) uses 0.64 GB; at $h = 12$ and $1{,}107$ parameters it uses 14.80 GB — $23\times$ the memory for $1.6\times$ the parameters — and at $h = 16$ and $1{,}603$ parameters it exhausts a 49 GB A6000 and cannot be trained at all. A SIREN of $1{,}675$ parameters reaches 40.35 dB on the same task in six seconds. The asymptotic obstruction comes from the architecture as specified rather than from this particular GPU, although the precise practical limit depends on the simulator and the differentiation strategy: width and qubit count are the same number, so the model cannot be made wider without leaving simulation.

5 Experimental Setup

For the QVF image study, all arms share the data, the latent-conditioned formulation, the optimiser (Adam), a common plateau scheduler using the released configuration's constants, and the epoch budget; they differ only in the network placed between coordinates and output.

Images. Fifty 32×32 CIFAR-10 (Krizhevsky, 2009) images from a single class are fitted simultaneously, each with a 64-dimensional latent code, following the released pipeline. We report PSNR of the fitted collection, with $\mathrm{PSNR} = 10 \log_{10}(1/\mathrm{MSE})$ on $[0, 1]$ data — the convention under which the original paper's reported MSE and PSNR are mutually consistent.

Arms. **QVF** is the authors' model with only the patches of §3. **QVF-frozen** is bit-identical except that the ansatz parameters do not receive gradients. **SIREN** (Sitzmann et al., 2020) uses the published initialisation with ω_0 swept. **FFN** is a positional-encoding ReLU network, the architecture used by NeRF (Mildenhall et al., 2020) and DeepSDF (Park et al., 2019). **Naive-Sin** uses periodic activations with default PyTorch initialisation and no frequency scaling, included to bracket what "MLP+Sin" can mean.

Table 1: Image collections: 50 CIFAR-10 images, 5000 epochs, one protocol. "+Sin" denotes the embedder activation the paper's best variant uses and the released code does not implement. Entries are single seed-0 runs; seeded estimates of the ansatz contribution are in Table 2 and differ. Grey rows are published values.

Model	Width	Params	PSNR (dB)	Time
SIREN (ω_0=30)	256	155,651	**41.14**	57 s
SIREN (ω_0=30)	192	92,163	38.19	51 s
SIREN (ω_0=30)	128	45,059	34.13	43 s
QVF+Sin, ansatz frozen	256	158,992	33.52	1,050 s
QVF+Sin	256	159,072	33.42	52,359 s
QVF+Sin	128	46,816	33.07	53,752 s
QVF+Sin, ansatz frozen	128	46,736	32.67	1,105 s
QVF, published (Gaussian+Sin)	—	—	32.2	—
FFN	256	155,651	30.71	43 s
QVF, published (Identity+ReLU)	—	—	29.5	—
QVF+ReLU, ansatz frozen	256	158,992	29.27	1,015 s
QVF+ReLU	256	159,072	28.78	75,960 s
QVF+ReLU	128	46,816	27.60	116,445 s
FFN	128	45,059	27.32	∼30 s
QVF+ReLU, ansatz frozen	128	46,736	27.22	1,885 s

Table 2: The ansatz ablation across five configurations. Only the gradient flow to the 80 quantum parameters differs between the columns.

Activation	Width	Frozen	Trained	Δ	95% CI on Δ
sin	128	33.11 ± 0.47	33.06 ± 0.18	-0.05	$[-1.54, +1.43]$
sin	192	32.86 ± 0.83	33.34 ± 0.79	$+0.47$	$[-1.17, +2.12]$
sin	256	33.16 ± 0.31	34.63 ± 1.01	$\mathbf{+1.48}$	$[\mathbf{+0.16}, \mathbf{+2.79}]$
ReLU	128	27.59 ± 0.32	27.60	$+0.01$	(one seed)
ReLU	256	28.70 ± 1.49	28.75 ± 2.31	$+0.05$	$[-9.24, +9.34]$

Mean $\pm$ sample standard deviation: five seeds at width 256 under the sin embedder, three elsewhere, except ReLU at width 128 (trained, one). The point estimate rises monotonically with width under the sin embedder. Δ is the mean paired per-seed difference and the interval is its Student 95% confidence interval; only the width-256 sin row excludes zero (paired t, $p = 0.036$, not corrected for the four comparisons).

6 IMAGES

We recover the published performance level. With the sin embedder, the variant the paper reports as its best and the release does not contain, our QVF reaches 33.07 dB at width 128 and 33.42 at width 256 against a published 32.2; the released ReLU variant reaches 28.78 at width 256 against a published 29.5. We take the model to be faithfully reconstructed and proceed to attribute its performance.

Almost all of the gain is classical. Between the two frozen-ansatz arms, where the circuit is the identity on both sides (Proposition 1), one activation change in the embedder moves the model from 27.59 ± 0.32 to 33.11 ± 0.47 dB, a gain of 5.52 dB. Training the 80 quantum parameters moves it by -0.05 dB at the same width and $+1.48$ dB at the largest we ran, at 40–75$\times$ the training time.

No stable gain at width 128, a measurable one at width 256. Table 2 repeats the freezing experiment across three widths and two activations. At width 128 the $+\sin$ model reaches 33.06 ± 0.18 dB trained against 33.11 ± 0.47 dB frozen: -0.05 dB, against a seed spread nine times larger. The paired interval is $[-1.54, +1.43]$, so what we can say is that this experiment rules out a gain larger than

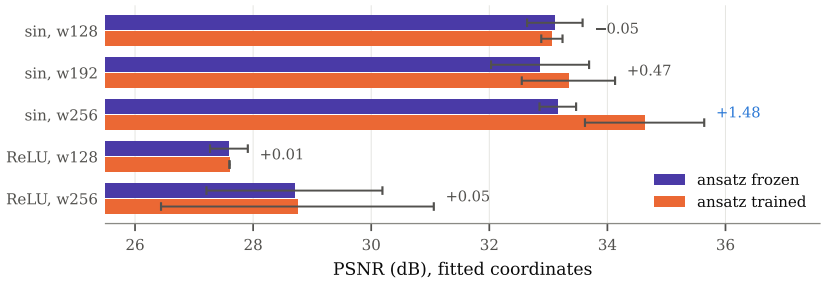

Figure 2: What the ansatz contributes. Freezing exactly the 80 quantum parameters, with architecture, initialisation, data and optimiser unchanged, leaves PSNR unchanged at width 128 and raises it by 1.48 dB at width 256, the only multi-seed configuration whose paired 95% interval excludes zero. Five seeds at width 256 under the sin embedder, three elsewhere, except ReLU at width 128 (trained). Training the ansatz costs 40–75× the wall-clock in every case.

about 1.4 dB and detects none; three seeds cannot establish that the contribution is negligible, which would require a pre-stated equivalence margin and a non-inferiority test we do not run. At width 256, where we ran five seeds, the same comparison gives 34.63 ± 1.01 against 33.16 ± 0.31: a paired gain of 1.48 dB with a 95% confidence interval of $[+0.16, +2.79]$, the largest effect we observe and the only one of the five whose interval excludes zero. The intermediate width sits between them at $+0.47$ dB, inside its interval, and both ReLU rows are null ($+0.01$, $+0.05$).

We report the monotone ordering as a trend we can see but have not established. Three widths at three to five seeds support a point estimate, not a slope, the width-256 interval is wide and close to zero at its lower end, and we have not corrected for testing five configurations. We flag it because it is the one place in this study where the ansatz does something measurable, and because an earlier draft of this paper — written before the width-256 arms were seeded — claimed the contribution was indistinguishable from zero everywhere. Two of our own single-seed readings did not survive repetition, a $+0.40$ dB gain and a -0.49 dB loss, both of which vanished once either side was repeated. Extending the width-256 arms from three seeds to five moved the estimate from $+0.83$ to $+1.48$ dB rather than towards zero, and two runs repeated under the final code reproduced bit-identically, so the spread across seeds is seed variance and not run-to-run noise.

The ablation is run under the stronger of the two initialisations. The released code offers a second initialisation, drawing the angles from $\mathcal{N}(0, 0.1)$ instead of the identity block. It is a different circuit, not merely a different starting point: the identity block stores each rotation twice, once forward and once adjoint, so it carries 80 parameters where the Gaussian branch carries 40. We repeated the width-256 ablation there. Trained, it reaches 32.89 ± 1.66 dB over five seeds against the identity block's 34.63 ± 1.01, a paired -1.74 dB with interval $[-3.72, +0.24]$; the configuration we audit is therefore not the weaker of the two, though at this sample size we cannot say it is the stronger. The frozen arm, however, is no longer a clean control. Proposition 1 does not apply, so freezing leaves a fixed *random* quantum map rather than the identity, and in 2 of 8 seeds that map was bad enough to cost roughly 14 dB (18.78 and 19.37 against 33.08 ± 0.63 for the other six). The resulting trained-minus-frozen difference, $+2.96$ dB with interval $[-2.90, +8.83]$, mostly measures recovery from a poor draw rather than any contribution of the ansatz. This is the reason we run the ablation under identity-block initialisation: it is the released default, it is not the weaker configuration, and it is the only one of the two for which freezing yields an exactly classical comparator.

A correctly initialised SIREN is ahead at both budgets. At width 128 SIREN reaches 34.29 ± 0.13 dB over five seeds against QVF+Sin's 33.06 ± 0.18; at width 256, 38.72 ± 1.98 against 34.63 ± 1.01. The margin grows with budget because the two families scale differently (§B). SIREN also reaches these values in under a minute, against 50,000 seconds and more.

The published classical baseline is consistent with a single hyperparameter setting. Sinusoidal INRs are governed by a first-layer frequency scale ω_0, for which the SIREN paper prescribes 30.

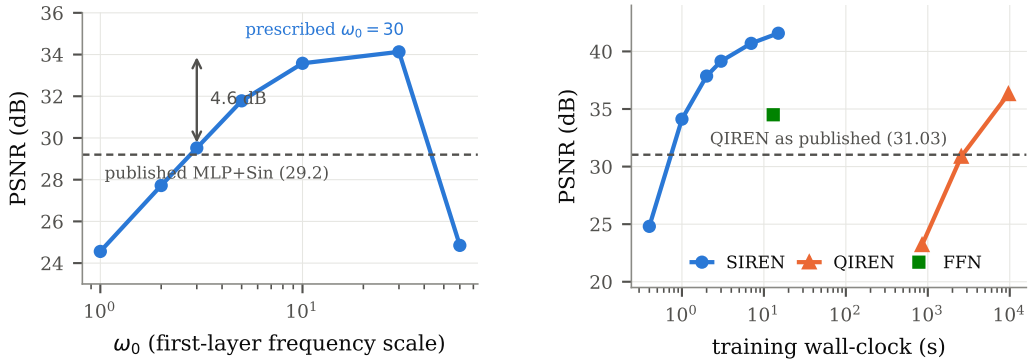

Figure 3: **Left:** the sinusoidal classical arm against its frequency scale, width 128. The optimum is the prescribed ω_0=30; the published MLP+Sin value matches what we measure an order of magnitude below it. **Right:** the single-image protocol at 675 parameters. Both arms improve with budget, three orders of magnitude of wall-clock apart; the dashed line is QIREN's published figure, reproduced to within 0.16 dB.

Sweeping it under our protocol (Figure 3, left; values in Table 7) produces a sharp optimum there and degrades quickly on either side. The paper's MLP+Sin entry, 29.2 dB, sits within 0.3 dB of what we measure at ω_0=3, an order of magnitude below the prescribed value and 4.6 dB below the optimum. We cannot verify their setting from the release, which contains no baseline implementation, so we report this as a consistency observation rather than an identification. It is not an implausible one: the reference implementation of the QINR that QVF compares against (Zhao et al., 2024) ships a training script whose default is ω_0=1, where we measure 24.56 dB.

Independently, because the identity-block initialisation makes the circuit an identity map, our frozen QVF+Sin arm is functionally the paper's own classical ablation; it reaches 32.67 dB where the paper reports 29.2 dB.

The constraint does not buy generalisation either. A constrained hypothesis set can pay off as a regulariser even when it costs fitting accuracy, and reconstruction error cannot see this. We therefore refit every arm on a random 75% of coordinates and report PSNR on the 25% never shown to the model (Table 3). Three things change and one does not.

The one that does not: freezing the ansatz is at least as good as training it, 26.12 ± 0.08 against 25.93 ± 0.09, the ranges again not overlapping. On fitted coordinates the ansatz's effect at this width was inside the seed spread; on held-out coordinates the ordering is consistent and in the same direction.

The circuit and its unconstrained relaxation are level (25.93 ± 0.09 against 25.87 ± 0.05; the budget-matched surrogate is level too, 25.80 ± 0.25). We flag that our own first reading of this row, taken from a single seed of the circuit, showed the circuit 0.10 dB ahead and led us to write that the restriction of Proposition 2 acts as a mild regulariser. A second seed removed the effect. We record the correction rather than the conclusion it briefly supported; it is the third time in this study that a single-seed reading did not survive repetition, which is itself the argument for reporting seeds.

What does change, and cuts against the reading we would otherwise have given: SIREN, which leads every fitting comparison in this paper by more than a decibel, is *last* here at 25.08 ± 0.50 — it overfits the observed coordinates hardest — and QVF's architecture as a whole therefore generalises about 1.0 dB better than a same-budget SIREN. That is a point in the original design's favour which a fitting-only comparison, including our own until we ran this, does not surface. It concerns the classical architecture; the ansatz is on the wrong side of the frozen-versus-trained comparison in this regime as well.

Convergence, not just endpoints. Figure 1(b) tracks SIREN and the released QVF variant against wall-clock at width 128; Table 4 gives the same run by epoch.

Table 3: Held-out coordinates: fit on a random 75% of pixels, evaluate on the rest. Width 128, three seeds per row; fitted values repeated for contrast.

Model	Fitted (dB)	Held-out (dB)	Drop
QVF+Sin, ansatz frozen	33.11 ± 0.47	$\mathbf{26.12 \pm 0.08}$	-6.99
QVF+Sin, ansatz trained	33.06 ± 0.18	25.93 ± 0.09	-7.13
Unconstrained M (classical)	33.83 ± 0.19	25.87 ± 0.05	-7.96
Budget-matched M (80 params)	33.85 ± 0.56	25.80 ± 0.25	-8.05
SIREN (ω_0=30)	$\mathbf{34.29 \pm 0.13}$	25.08 ± 0.50	-9.21

Table 4: Convergence at width 128, 50 images, released (ReLU) QVF variant.

Epochs	SIREN (dB)	Time	QVF+ReLU (dB)	Time
200	24.01	2 s	—	—
400	27.50	3 s	—	—
1000	30.79	7 s	25.73	10,553 s
2000	32.73	17 s	27.09	21,848 s
5000	34.13	31 s	27.60	116,445 s

7 A Second QINR: QIREN in the Same Harness

QVF's other point of comparison is QIREN (Zhao et al., 2024), an earlier QINR that fits a single signal rather than a conditioned collection, so adding it requires a single-image protocol and two compatibility patches (§F). It places one qubit per hidden unit: width 8 gives 675 parameters on 8 qubits, width 16 gives 1,603 on 16 qubits and already exhausts the 49 GB accelerator we used (§4). Its budget is capped by simulability, and we could not train it beyond about a thousand parameters; the 0.74×10^5 parameters QVF reports for it would need of order a hundred qubits, a figure we could not reconcile. Under its own prescribed settings, learning rate 0.03 with a cosine schedule, QIREN reaches 30.87 dB at 675 parameters against the 31.03 reported in QVF's Table 2. Those settings matter: under the classical arms' learning rate it reaches 20.12 dB, so an untuned comparison would have been unfair by 10.75 dB, and we swept the rate for both families (Table 6). Tuned, QIREN does close on the classical arm, 36.31 dB at 1000 epochs against SIREN's 37.86, though the same comparison read 3.25 dB at 300 epochs, which is why we report the curve. The cost does not close: that 1000-epoch value takes SIREN two seconds and QIREN 9,628, and fifteen seconds, still $1/640$ of QIREN's budget, puts SIREN 5.27 dB ahead at 41.58 (Figure 3, right). At 675 parameters the two are close in what they can represent and three to four orders of magnitude apart in what it costs, the same shape as for QVF.

8 Conclusion

We rebuilt two quantum implicit neural representations under a common evaluation harness at matched parameter budgets, recovered the reported performance level for both, and asked what produces it. For QVF the answer is almost entirely classical: one activation change in the embedder is worth 5.52 dB, while training the 80-parameter ansatz, 0.17% of the model, is worth -0.05 dB at width 128 and $+1.48$ dB at width 256, the only multi-seed configuration whose paired 95% confidence interval excludes zero. A proposition shows the circuit's output to be a quadratic form in the embedder's own output, selected from a family of intrinsic dimension at most 80; a classical readout holding the same 80 parameters learns that form three orders of magnitude faster and, at width 256, 2.85 dB better. The proposition bounds the readout, not the embedder, so it does not by itself forbid a wider classical network from helping; empirically the architecture is nonetheless flat in classical width, and its one effective scaling direction is the one whose simulation cost is exponential (§B).

Reproducibility Statement

Our harness, the patches described in §3, the pinned environment specification, and the raw output of every run reported here are released with this submission. Each table entry corresponds to a single logged run; wall-clock figures are from one NVIDIA A6000 on a shared machine and should be read as ratios rather than absolute times.

References

Reyhaneh Aghaei Saem, Behrang Tafreshi, Zoë Holmes, and Supanut Thanasilp. Pitfalls when tackling the exponential concentration of parameterized quantum models. *Quantum Science and Technology*, 11(1):015049, 2026.

Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Nathan Killoran, et al. PennyLane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2018.

M. Cerezo, Martin Larocca, Diego García-Martín, N. L. Diaz, Paolo Braccia, Enrico Fontana, Manuel S. Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, Eric R. Anschuetz, and Zoë Holmes. Does provable absence of barren plateaus imply classical simulability? *Nature Communications*, 16:7907, 2025. doi: 10.1038/s41467-025-63099-6.

Paolo Cremonesi and Dietmar Jannach. Progress in recommender systems research: Crisis? what crisis? *AI Magazine*, 42(3), 2021.

Amer Essakine, Yanqi Cheng, Chun-Wun Cheng, Lipei Zhang, Zhongying Deng, Lei Zhu, Carola-Bibiane Schönlieb, and Angelica I. Aviles-Rivero. Where do we stand with implicit neural representations? a technical and performance survey. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=QTsJXSvAI2.

Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. Are we really making much progress? a worrying analysis of recent neural recommendation approaches. In *ACM Conference on Recommender Systems (RecSys)*, 2019.

Elies Gil-Fuster, Jonas R. Naujoks, Grégoire Montavon, Thomas Wiegand, Wojciech Samek, and Jens Eisert. Opportunities and limitations of explaining quantum machine learning. *arXiv preprint arXiv:2412.14753*, 2024.

Elies Gil-Fuster, Casper Gyurik, Adrián Pérez-Salinas, and Vedran Dunjko. On the relation between trainability and dequantization of variational quantum learning models. In *International Conference on Learning Representations (ICLR)*, pp. 24069–24093, 2025.

Syeda Anshrah Gillani, Mirza Samad Ahmed Baig, Shahid Munir Shah, Asher Ali, and Hamzah Siddiqui. How quantum is the advantage? a fair, calibration- and noise-aware benchmark and attribution audit of quantum machine learning for network intrusion detection. *arXiv preprint arXiv:2608.18155*, 2026.

Cyrille Yetuyetu Kesiku and Begonya Garcia-Zapirain. How quantum circuits actually learn: A causal identification of genuine quantum contributions. *arXiv preprint arXiv:2603.16321*, 2026.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

David B. Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. BACON: Band-limited coordinate networks for multiscale scene representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao. FINER: Flexible spectral-bias tuning in implicit neural representation by variable-periodic activation functions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1): 4812, 2018.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.

Alireza Morsali, MohammadJavad Vaez, Mohammad Hossein Soltani, Amirhossein Kazerouni, Babak Taati, and Morteza Mohammad-Noori. A unified theory of sinusoidal activation families for implicit neural representations. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=ZDmBPYptbL.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4), 2022.

Sreeraj Rajindran Nair and Christopher Ferrie. Local surrogates for quantum machine learning. *arXiv preprint arXiv:2506.09425*, 2025.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Sameera Ramasinghe and Simon Lucey. Beyond periodicity: Towards a unifying framework for activations in coordinate-MLPs. In *European Conference on Computer Vision (ECCV)*, 2022.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G. Baraniuk. WIRE: Wavelet implicit neural representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Franz J. Schreiber, Jens Eisert, and Johannes Jakob Meyer. Classical surrogates for quantum learning models. *Physical Review Letters*, 131(10):100803, 2023.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum machine learning models. *Physical Review A*, 103(3):032430, 2021.

Harald Semmelrock, Tony Ross-Hellauer, Simone Kopeinik, Dieter Theiler, Armin Haberl, Stefan Thalmann, and Dominik Kowald. Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine*, 46(2):e70002, 2025. doi: 10.1002/aaai.70002.

Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Shuteng Wang, Christian Theobalt, and Vladislav Golyanik. Quantum visual fields with neural amplitude encoding. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Jiaming Zhao, Wenbo Qiao, Peng Zhang, and Hui Gao. Quantum implicit neural representations. In *International Conference on Machine Learning (ICML)*, 2024.

A AI USE STATEMENT

Parts of this paper's text, code and analysis were produced with the assistance of generative AI. The authors take full responsibility for the final content of this paper, including any text produced with the assistance of generative AI.

B Scaling Behaviour

Three knobs can be turned, the embedder's width, the circuit's depth and the qubit count, and they behave very differently. With the ansatz frozen, QVF measures 33.11, 32.86 and 33.16 dB at widths 128, 192 and 256: flat to within seed noise across a doubling of the classical network, where a SIREN over the same range goes $34.29 \rightarrow 37.83 \rightarrow 38.72$ dB. Doubling the circuit's repetitions from $J{=}2$ to $J{=}4$ moves PSNR by $+0.01$ dB ($29.15 \rightarrow 29.16$ at 1000 epochs) for twice the wall-clock. Only the qubit count moves the model: raising $d = 2^n$ from 16 to 64 improves it by 1.26 dB ($29.15 \rightarrow 30.41$). Even that gain is mostly not the circuit's: the same change with the ansatz frozen at identity yields $+1.01$ dB ($28.28 \pm 1.73 \rightarrow 29.29 \pm 0.60$), because enlarging d also widens the embedder's output layer by 13% of the model's parameters, leaving about 0.25 dB attributable to the extra qubits and well inside the seed spread.

The observation is consistent with the bottleneck identified by Proposition 2: the output is $\psi^\top M \psi$ with ψ on the simplex Δ^{d-1}, so everything the embedder computes is read out through d numbers fixed by the qubit count and not by the width. This does not prove that a wider embedder cannot help, since it could still learn a more useful $x \mapsto \psi(x)$ within the same d; it does say where the ceiling would come from, and widening d is in fact the one intervention that moves the model. That intervention is also the expensive one: $d = 2^n$, so its simulation cost is exponential in the number of qubits, and Proposition 3 makes the same point for QIREN, where width and qubit count are the same number and a 1,603-parameter model exhausts the 49 GB accelerator we used. Both architectures are limited in the same direction, and the only effective scaling knob is the one that cannot be scaled.

C What We Could Not Establish

Scope. We evaluate two QINRs on images at the scales their authors chose. We also ran the shape modality and exclude it: our runs matched neither the published absolute values nor the published ordering of the two embedder activations, and we describe that failure in §C rather than drawing conclusions from a reconstruction we cannot vouch for. Nothing here bears on whether quantum circuits may become useful in INRs at scales or on hardware we cannot access; our claims concern what is measurable today, in simulation, at the sizes reported.

Reproduction, and what we excluded. We reproduce the paper's best reported image result (33.06 ± 0.18 against 32.2 dB) and its released variant to within 0.7 dB (28.78 against 29.5). We also ran the shape pipeline and do not report it. Our runs there matched neither the published absolute values nor, more seriously, the published ordering of the two embedder activations, which we measured in the opposite direction. The repository ships three of the five shapes its configuration requests, which may account for some of the gap but not for a reversal. Rather than draw conclusions from a reconstruction we cannot vouch for, we restrict every claim in this paper to images, where the reproduction succeeds. The shape code is still covered by the audit of §3, which concerns the released artifact rather than any result obtained from it.

The baselines we criticise may not be the baselines that were run. The original paper describes its comparators both as "widely-used classical foundational INR methods, i.e., MLPs with ReLUs, and Siren" and as differing from QVF only in "the presence of the ansatz". These readings are not compatible: the second describes ablations of the authors' own architecture, the first describes independent methods. Our 34.13 dB is a properly initialised SIREN and is therefore a fair comparator under the first reading only. Under the second, our contribution is narrower and methodological: an ablation of one's own architecture is not a substitute for the field's baseline, and the two should not be reported under the same name. We were not able to determine which of the two was run.

D Discussion

Having reconstructed the model well enough to reproduce its published number, we can decompose it. Between the two frozen-ansatz arms, where the circuit is the identity on both sides, replacing one activation function in the classical embedder is worth 5.52 dB ($27.59 \pm 0.32 \rightarrow 33.11 \pm 0.47$).

Training the ansatz on top of that is worth -0.05 dB at width 128 and $+1.48$ dB at width 256. The circuit is 0.17% of the parameters, and the classical change outweighs its best showing by more than three to one and its typical showing entirely. We report only directly measured quantities; nothing in this decomposition is extrapolated.

We want to be careful about what this does and does not show. It does not show that quantum circuits are useless in INRs; we measure a small effect, not the absence of one, and a different construction could carry more. It does not show that the original results are wrong; we reproduce them. What it shows is narrower and, we think, still worth stating: under this reconstruction and this matched protocol, the available evidence does not support attributing the reported gain primarily to the quantum ansatz. A classical activation change accounts for $3.7\times$ as much as the ansatz's largest observed gain, and substantially more than its effect in the other configurations. Whether a different implementation or protocol would restore the attribution we cannot say; what we can say is that a single frozen-parameter run would have surfaced the question.

The pattern is not specific to quantum machine learning. Fields that have audited themselves have repeatedly found reported gains resting on baselines given less care than the proposed method (Ferrari Dacrema et al., 2019; Cremonesi & Jannach, 2021), and the methodological remedies are by now well-catalogued (Semmelrock et al., 2025). What is specific here is that the quantum component makes the asymmetry easy to miss: it is the part that looks novel, so it is the part that gets credited. Calibration- and noise-aware attribution audits in other quantum application areas (Gillani et al., 2026) reach a similar conclusion by a different route.

The procedural recommendation follows directly. QINR papers should report the parameter split between quantum and classical components, should include a frozen-ansatz ablation, should ensure the released code implements the variant whose numbers are reported, and should distinguish ablations of their own architecture from independent implementations of the classical methods they name. Each is inexpensive. Together, these checks would make attribution substantially more transparent and reduce the need for post-hoc audits.

E Derivations and Numerical Verification

Proposition 1. Identity-block initialisation sets the rotation and entangling parameters so that $w_{1:n} = w_{n+1:2n}$ and $c_{1:n} = c_{n+1:2n}$. Each repetition of the circuit applies, in order, the entangler $E(c_{1:n})$, the rotations $R(w_{1:n})$, the adjoint rotations $R(w_{n+1:2n})^\dagger$ and the adjoint entangler $E(c_{n+1:2n})^\dagger$. Under the equalities above the second pair inverts the first, so each repetition is $E(c)R(w)R(w)^\dagger E(c)^\dagger = \mathbb{I}$ and the product over repetitions is $\mathbb{I}$. The readout then reduces to $f_c(x) = \sum_i P_{ci}\big(1 - 2\Pr[\text{bit } i = 1]\big)$, a function of $p(x)$ alone. Numerically, applying the frozen layer to the computational basis reproduces the basis-state parities to 2.4×10^{-7} in single precision.

Proposition 2. The ansatz is built from R_Y and CR_Y gates only. Both have real matrices, so $U(\boldsymbol{\theta}_q)$ is real orthogonal, not merely unitary, and $M_c = \sum_i P_{ci} U^\top Z_i U$ is real symmetric rather than merely Hermitian. The embedder emits $\psi(x) = \sqrt{p(x)}$, which is real and non-negative, so

$$\langle Z_i \rangle = \psi(x)^\top U^\top Z_i U\, \psi(x), \qquad f_c(x) = \psi(x)^\top M_c\, \psi(x),$$

with M_c independent of x. Because $U(\boldsymbol{\theta}_q)$ is a product of R_Y and CR_Y gates whose entries are trigonometric functions of the angles, it is real-analytic in $\boldsymbol{\theta}_q$, and so is $\boldsymbol{\theta}_q \mapsto (M_1, \ldots, M_C)$. The differential of a smooth map from a $\dim \boldsymbol{\theta}_q$-dimensional space has rank at most $\dim \boldsymbol{\theta}_q$ everywhere, so its image has dimension at most $\dim \boldsymbol{\theta}_q$. (Continuity alone would not suffice: a space-filling curve is a continuous surjection onto a square.) The tuple lies in $(\mathrm{Sym}_d)^C$, of ambient dimension $C\binom{d+1}{2}$; each individual M_c lies in a $\binom{d+1}{2}$-dimensional space.

We verify the two structural facts numerically. Building U column by column from the circuit's own gate sequence gives $\max|\operatorname{Im} U| = 0$ and $\|U^\top U - I\|_\infty = 6.7 \times 10^{-16}$, and the resulting M_c is exactly symmetric. Probing the layer with vectors of the form $(e_a + e_b)/\sqrt{2}$ to recover M_c entrywise, and comparing $\psi^\top M_c \psi$ against the circuit's own output on 64 embedder-shaped inputs, agrees to 2.4×10^{-7}.

We also verify the readout matrix and the composed map, since the released parity routine is easy to mis-state. Probing it with the standard basis returns P exactly as displayed in §4: it sums rather

than averages, and the padding sends z_3 into the red channel. Carrying this through, the model's RGB output agrees with $\psi^\top M_c \psi$ to 2.4×10^{-7} and $\partial f_R / \partial \psi$ agrees with $2M_R\psi$ to 3.6×10^{-7}, so the quadratic-form description holds for the quantities that are actually trained and not only for the intermediate measurements. Since $\operatorname{rank} P = 3$ while $n = 4$, the output is invariant to any change of M_0 and M_3 that preserves $M_0 + M_3$; the effective ambient dimension is therefore $3\binom{d+1}{2} = 408$ rather than 544.

Were the gate set complex, the same argument would give $f_c = \psi^\top \operatorname{Re}(M_c)\,\psi$, with the reachable set of $\operatorname{Re}(M_c)$ in place of M_c; the dimension counting is unchanged.

Proposition 3. A QIREN hybrid layer of width h applies a classical linear map into h dimensions, normalises, and evolves an h-qubit state, whose amplitude vector has 2^h entries; holding it requires $\Theta(2^h)$ memory, and applying the $\operatorname{poly}(h)$ gates of a layer costs $\operatorname{poly}(h)\,2^h$ time per sample. The trainable parameters are the linear map ($\Theta(h^2)$), the batch-norm statistics ($\Theta(h)$) and the circuit's rotation angles ($\Theta(h)$ per entangling layer), so $P = \Theta(h^2)$ and $h = \Theta(\sqrt{P})$. Substituting gives $\Theta(2^{\sqrt{P}})$ memory and $\operatorname{poly}(\sqrt{P})\,2^{\Theta(\sqrt{P})}$ time.

F Experimental Details

F.1 Supporting tables

Table 5: The surrogate of Proposition 2: an unconstrained classical superset of the circuit's function class, not the same class, together with a variant restricted to the ansatz's own 80-parameter budget. All rows use the sin embedder except where noted; 50 images, 5000 epochs, three seeds where an interval is given. Classical readout timings were measured together on an idle device, separately at each width.

Readout	**Width**	**Params**	**PSNR (dB)**	**Time**
Unconstrained M (classical)	128	47,760	33.83 ± 0.19	60 s
Budget-matched M (80 params)	128	46,816	$\mathbf{33.85 \pm 0.56}$	62 s
Quantum circuit	128	46,816	33.06 ± 0.18	53,752 s
Circuit frozen ($= \mathbb{I}$)	128	46,736	33.11 ± 0.47	1,105 s
Unconstrained M (classical)	256	160,016	36.47 ± 0.55	102 s
Budget-matched M (80 params)	256	159,072	$\mathbf{36.79 \pm 0.46}$	125 s
Quantum circuit	256	159,072	34.63 ± 1.01	52,290 s
Circuit frozen ($= \mathbb{I}$)	256	158,992	33.16 ± 0.31	1,868 s
Quantum circuit, ReLU embedder	128	46,816	27.60	116,445 s
Unconstrained M, ReLU embedder	128	47,760	27.55 ± 0.55	59 s

Table 6: Single-image protocol at 675 parameters. Learning rate swept for both; each method is shown at its own best setting. QIREN's published figure is 31.03 dB.

Epochs	**SIREN (dB)**	**Time**	**QIREN (dB)**	**Time**
100	24.81	< 1 s	23.20	851 s
300	34.12	1 s	30.87	2,560 s
1000	37.86	2 s	36.31	9,628 s
2000	39.15	3 s	—	—
5000	40.70	7 s	—	—
10000	**41.58**	15 s	—	—
FFN reference: 34.50 *dB at* 10000 *epochs,* 13 *s*				

F.2 Environment

All runs use the environment pinned by the authors' `environment.yaml`: Python 3.10.12, PennyLane 0.34.0 with `default.qubit.torch`, PyTorch 2.1.2+cu121, torchvision 0.16.2, NumPy 1.26.3. Hardware is a single NVIDIA RTX A6000 (49 GB) on a shared, contended host; wall-clock figures should therefore be read as ratios. We verified that the same code fails on a newer PennyLane (0.42), where the QNode return convention changed, and we did not attempt to run the authors' model outside its pinned versions.

F.3 Patches applied to the released code

We changed four lines in total. The image pipeline required two:

```
- self.pixels = images.view(num_samples, h * w, 3)
+ self.pixels = self.images.reshape(num_samples, h * w, 3)

- self._rotations(weights[:self.n_qubits, :, r])
- qml.adjoint(self._rotations)(weights[self.n_qubits:, :, r])
+ self._rotations(weights[:self.n_qubits, r])
+ qml.adjoint(self._rotations)(weights[self.n_qubits:, r])

- self._rotations(weights[:, :, r])
+ self._rotations(weights[:, r])
```

The first restores the truncation the dataset performs one line earlier. The remaining three index a rank-2 parameter tensor with two subscripts instead of three; the replacement is character-for-character the indexing used by the authors' own shape pipeline, which we take as evidence that it restores the intended semantics rather than substituting a different circuit. Under it the identity-block initialisation behaves as designed, with the forward and adjoint rotations cancelling at initialisation.

The shape pipeline required a relative data path and one configuration key, `parity_channels`, which the code reads but the shipped configuration does not define; we set it to 1, matching the `output_channels` value the configuration does provide. We made no other change to either model.

F.4 Protocol

Every arm — the authors' and ours — is trained by the same loop. Coordinates receive the released positional encoding (L=6, input included) and are concatenated with a per-signal latent code that is optimised jointly by a separate Adam instance, as in the released code. The loss is the released objective, mean squared error plus the same latent-norm regulariser at σ^2=10^{-4}. We add a `ReduceLROnPlateau` scheduler with the released configuration's constants (factor 0.9, patience 15); the original paper's text instead specifies a window of 50 epochs, a discrepancy we discuss in §3. It is applied identically to all arms. Learning rates are the released values (10^{-2} for both model and latents), except where an arm's own prescribed setting differs and is swept (§7).

F.5 Arm definitions

QVF is `ImageRegressor` / `GeoRegressor` unmodified apart from the patches above. **QVF-frozen** is the same object with `requires_grad_(False)` applied to the ansatz parameters and nothing else; the circuit still executes in the forward pass. **SIREN** uses $\mathcal{U}(-1/d_{\text{in}}, 1/d_{\text{in}})$ on the first layer and $\mathcal{U}(-\sqrt{6/d_{\text{in}}}/\omega_0, \sqrt{6/d_{\text{in}}}/\omega_0)$ thereafter, with $\sin(\omega_0 \cdot)$ activations and zero biases. **FFN** is the same depth and width with ReLU and default PyTorch initialisation. **Naive-Sin** is **FFN** with sin substituted for ReLU and no frequency scaling, i.e. ω_0=1 with default initialisation. **QVF+Sin** is the authors' model with the single hard-coded `F.relu` call in the embedder replaced by `torch.sin`, everything else untouched; this reconstructs the "+Sin" variant their tables report and their code omits. Because the identity-block initialisation makes the circuit an identity map, **QVF+Sin, frozen** is functionally their classical MLP+Sin ablation.

We also repeated the width-128 frozen ReLU run after correcting a wiring defect in our own harness — the `--hidden` flag initially set the width of the classical arms only, while the quantum arms read it from a configuration file we had edited. The repeat reproduces the original value exactly (27.22 dB), confirming that no quantum-arm result was affected.

Parameter matching is on trainable network parameters, excluding latent codes, which are identical across arms (50×64). At width 128 the classical arms hold 45,059 against QVF's 46,816 (3.7% fewer); at width 256, 155,651 against 159,072 (2.1% fewer). The classical arms are therefore never advantaged on budget.

F.6 Sweeps referenced from the main text

Table 7: Frequency-scale sweep for the sinusoidal classical arm, width 128. The prescribed value is ω_0=30; the published MLP+Sin entry (29.2 dB) matches our ω_0=3 measurement.

ω_0	1	2	3	5	10	30	60
PSNR (dB)	24.56	27.72	29.52	31.78	33.58	**34.13**	24.85

F.7 Complete results

Table 8 lists every run behind the figures, including configurations omitted from the main text.

F.8 Configuration keys the released code does not read

The image configuration defines eight keys no code path consumes: `M_energy_decay`, `M_quantum_decay`, `gaussian_initialization`, `latent_pretrained`, `lr_model_quantum`, `lr_multiplier`, `patience`, `pretrained`. The shape configuration defines eleven, additionally including `gaussian_std`, `infer_epochs`, `infer_samples_use`, `output_channels`, `train`, and `use_identity_block`. Two consequences are worth noting. Setting `gaussian_initialization` has no effect on either model, so the initialisation branch is selected solely by `use_identity_block` for images and is not selectable at all for shapes. And the pair (`lr_multiplier`, `patience`) matches the scheduler the paper describes, which suggests the scheduler was present in a version that produced the reported numbers and was lost before release.

F.9 On the widths we could not disambiguate

The original paper's appendix reports 0.52×10^5 parameters "corresponding to p=128" in a subsection comparing QVF with QIREN, which concerns its Table 2. Its released configuration sets width 256 ($\approx 1.59 \times 10^5$ parameters). Its Table 3, the source of the 32.2 dB entry we compare against, does not state which was used. We ran the main comparison at width 128 to match the stated parameter count, and report the width-256 arms alongside so that the comparison can be read either way. Under the first reading our matched SIREN gives 34.13 dB; under the second, 41.14 dB.

Table 8: Every run behind the figures and tables. Images: 50 CIFAR-10 images, 5000 epochs, PSNR in dB. Seed 0 unless noted. "+Sin" is the embedder activation the release does not implement.

Arm	Configuration	Params	Metric	Time (s)
Images — classical arms				
SIREN	ω_0=30, width 256	155,651	41.14	57
SIREN	ω_0=30, width 192	92,163	38.19	51
SIREN	ω_0=30, width 128	45,059	34.13	43
SIREN	ω_0=10, width 128	45,059	33.58	43
SIREN	ω_0=60, width 128	45,059	24.85	43
FFN	width 256	155,651	30.71	43
FFN	width 128	45,059	27.32	26
Naive-Sin	width 128, seed 0	45,059	22.73	28
Naive-Sin	width 128, seed 1	45,059	22.37	28
Naive-Sin	width 128, seed 2	45,059	23.80	28
Images — QVF, $+\sin$ variant (the paper's best)				
QVF+Sin, frozen	width 256	158,992	33.52	1,050
QVF+Sin	width 256	159,072	33.42	52,359
QVF+Sin	width 128	46,816	33.07	53,752
QVF+Sin, frozen	width 128	46,736	32.67	1,105
Images — QVF, ReLU variant (as released)				
QVF+ReLU, frozen	width 256	158,992	29.27	1,015
QVF+ReLU	width 256	159,072	28.78	75,960
QVF+ReLU	width 128	46,816	27.60	116,445
QVF+ReLU, frozen	width 128	46,736	27.22	1,885
QVF+ReLU, frozen	width 128, repeat run	46,736	27.22	6,932
Images — convergence, width 128, ReLU variant				
SIREN	200 epochs	45,059	24.01	2
SIREN	400 epochs	45,059	27.50	3
SIREN	1000 epochs	45,059	30.79	7
SIREN	2000 epochs	45,059	32.73	17
QVF+ReLU	1000 epochs	46,816	25.73	10,553
QVF+ReLU	2000 epochs	46,816	27.09	21,848
Images — Gaussian initialisation, width 256 (40-parameter circuit)				
Gaussian, trained	seeds 0–2	159,032	32.74, 34.24, 30.14	27,232–27,665
Gaussian, trained	seeds 3–4	159,032	34.15, 33.18	36,060–36,903
Gaussian, frozen	seeds 0–3	158,992	32.40, 32.84, *18.78*, 32.72	606–5,169
Gaussian, frozen	seeds 4–7	158,992	32.89, 34.14, *19.37*, 33.51	2,431–11,926