SMLab Talk

# Quantum Implicit Neural Representations

*Jiaming Zhao | Wenbo Qiao | Peng Zhang | Hui Gao*

# OUTLINES

- Introduction
- What Are Implicit Neural Representations (INRs)?
- Limitations of Classical INRs
- QIREN : A Quantum Solution
- Architecture of QIREN
- Why is it better?
- Datasets and Evaluations
- Conclusion & Future Work
- References

# INTRODUCTION

- **Resolution Dependency**: Traditional neural networks lose details when zoomed in, causing blurriness.
- **INRs Advantage**: INRs use continuous functions for flexible, high-quality reconstructions.
- **High-Frequency Challenge**: Classical INRs struggle to capture fine details with standard activations like ReLU.
- **FNNs Tradeoff**: Fourier Neural Networks improve performance but require high parameter counts.
- **QIREN Solution**: QIREN leverages quantum circuits for efficient and accurate signal representation.

# What Are Implicit Neural Representations (INRs)?

- **Traditional vs. INR**: Traditional methods store discrete data points, while INRs represent signals as continuous functions.
- **Coordinate-Based Mapping**: INRs map coordinates (e.g., (x, y)) to values (e.g., color, intensity), enabling smooth representations.
- **Resolution Independence**: Unlike grid-based methods, INRs allow seamless scaling and high-detail reconstructions.
- **Applications**: Used in **computer vision, graphics, and AI** for tasks like image synthesis, super-resolution, and 3D modeling.

# ReLU with Random Fourier Features

**A perception with RFF has the following form:**

$$\boxed{g(x) = W \cdot \begin{pmatrix} \cos(2\pi M \cdot x) \\ \sin(2\pi M \cdot x) \end{pmatrix} + b}$$

***Notation:***

$W \in R^{1\times 2m} \rightarrow$ Weight matrix

$M \in R^{m\times d_{in}} \rightarrow$ Random Fourier mapping matrix

$x \in R^{d_{in}} \rightarrow$ Input

$b \in R \rightarrow$ Bias

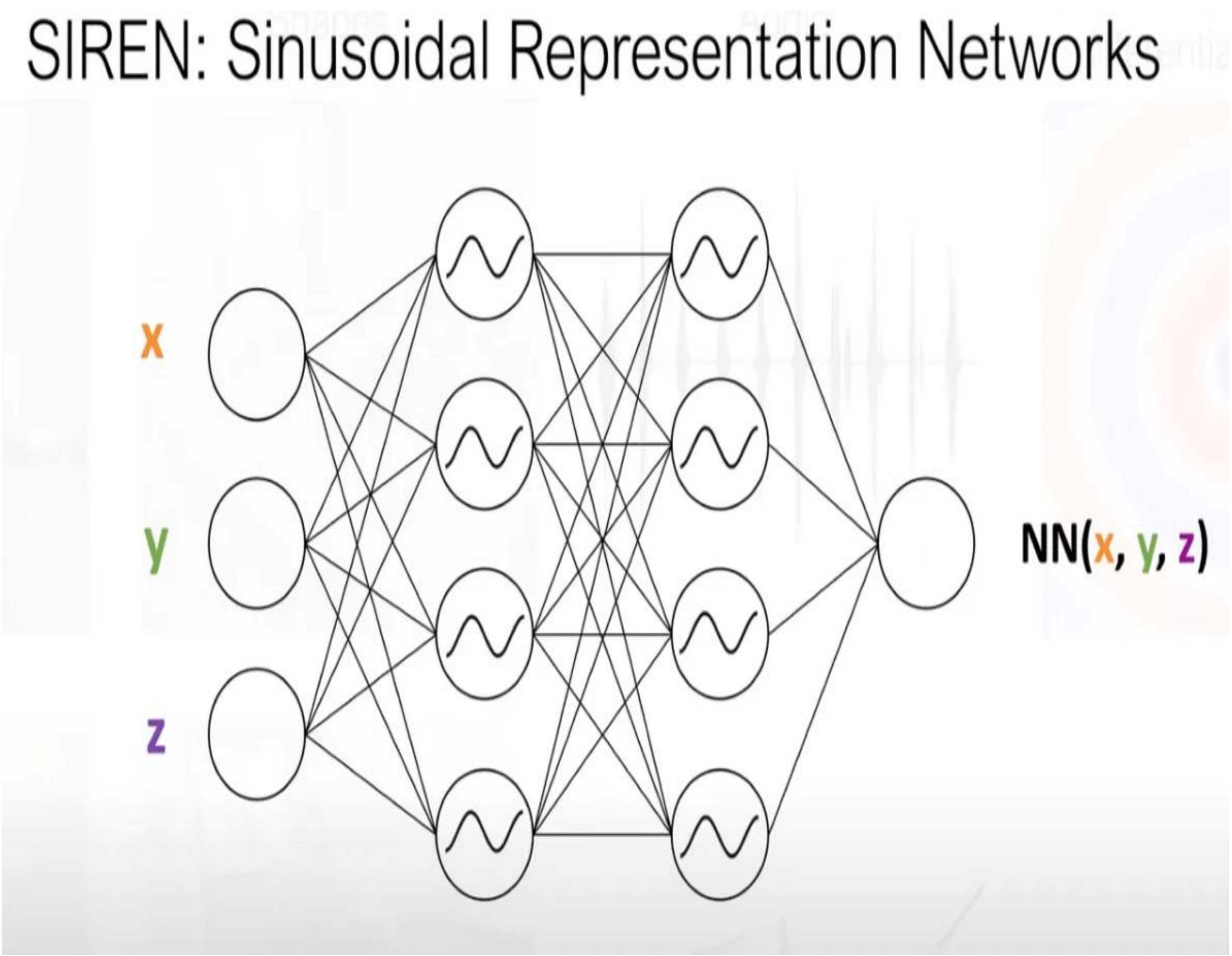
# SIREN

**SIREN is a MLP that utilizes the sine activation function. Considering SIREN with one hidden layer, we get:**

$$\boxed{g(x)=W_2 \cdot \sin(\,2\pi W_1 \cdot\, x + \theta\,) + b}$$

Notation:

- $W_1 \in \mathbb{R}^{2m \times d_{in}}$ → Parameter matrix
- $W_2 \in \mathbb{R}^{1 \times 2m}$ → Parameter matrix
- $\theta \in \mathbb{R}^{2m}$ → Bias vector
- $x \in \mathbb{R}^{d_{in}}$ → Input
- $b \in \mathbb{R}$ → Bias




***Source:*** https://medium.com/@aryamansriram/paper-review-implicit-neural-representations-with-periodic-activation-functions-c1e00179cb8e

# Limitations of Classical INRs

- **High-Frequency Limitations**: ReLU-based models miss fine textures and sharp edges.
- **Spectral Bias**: Classical models prefer low frequencies, failing on complex details.
- **Parameter Inefficiency**: FNNs need many parameters to capture details.
- **Computational Overhead**: Larger models increase memory use and computation costs.

# QIREN

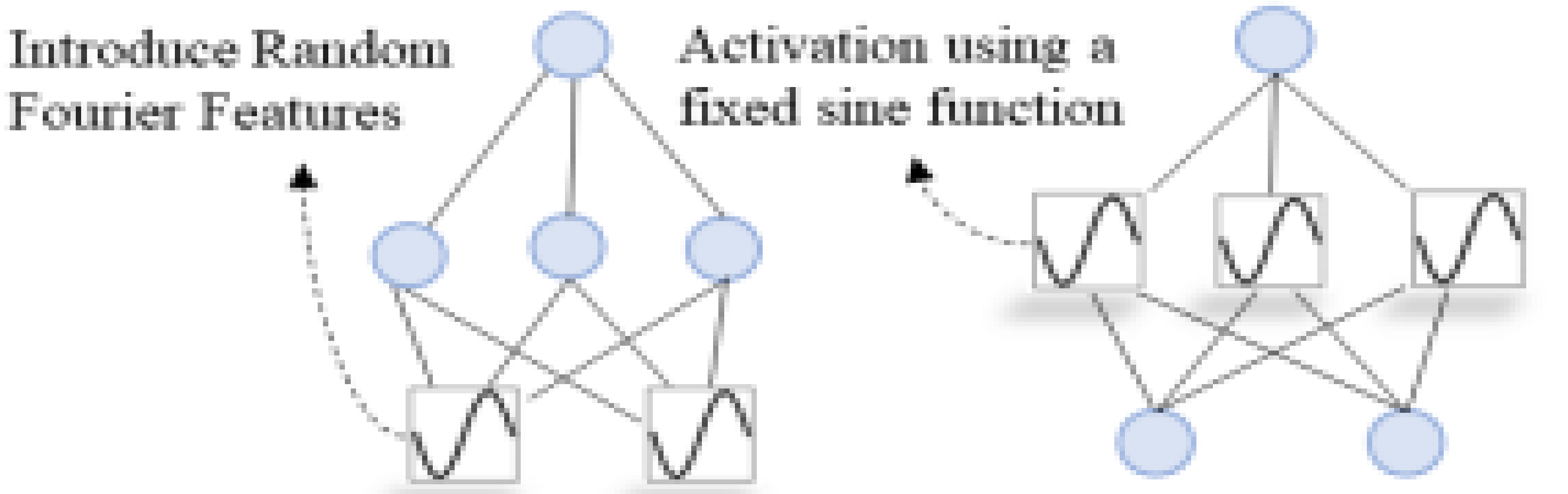
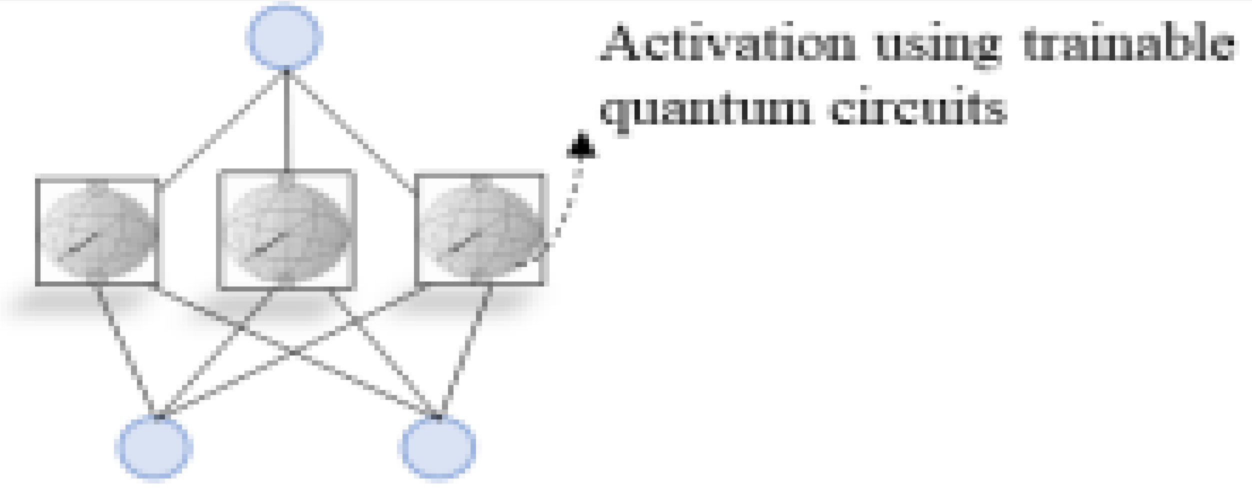
| | | |
|---|---|---|
| **Domain** | Implicit Neural Representations (INRs) | |
| **Challenge** | Modeling high-frequency components of signals | |
| **Methodology** | Fourier Neural Networks (FNNs) | |
| **Model** | Classical Implicit Representation Network | Quantum Implicit Representation Network (QIREN) |
| |   (a) ReLU-based MLPs with Random Fourier Features (RFF) (b) Sinusoidal Representation Network (SIREN) |   (c) A data re-uploading circuit is introduced to realize the quantum generalization of FNN. |
| **Signal representation capacity** | Growing **linearly** with the model size | Growing **exponentially** with the size of quantum circuits under optimal conditions |
| **Result** | More parameters and less precise signal representation | **Fewer parameters and more precise signal representation** |

*Figure 2.* Classical Fourier Neural Networks vs. Quantum Fourier Neural Networks.

# ARCHITECTURE OF QIREN

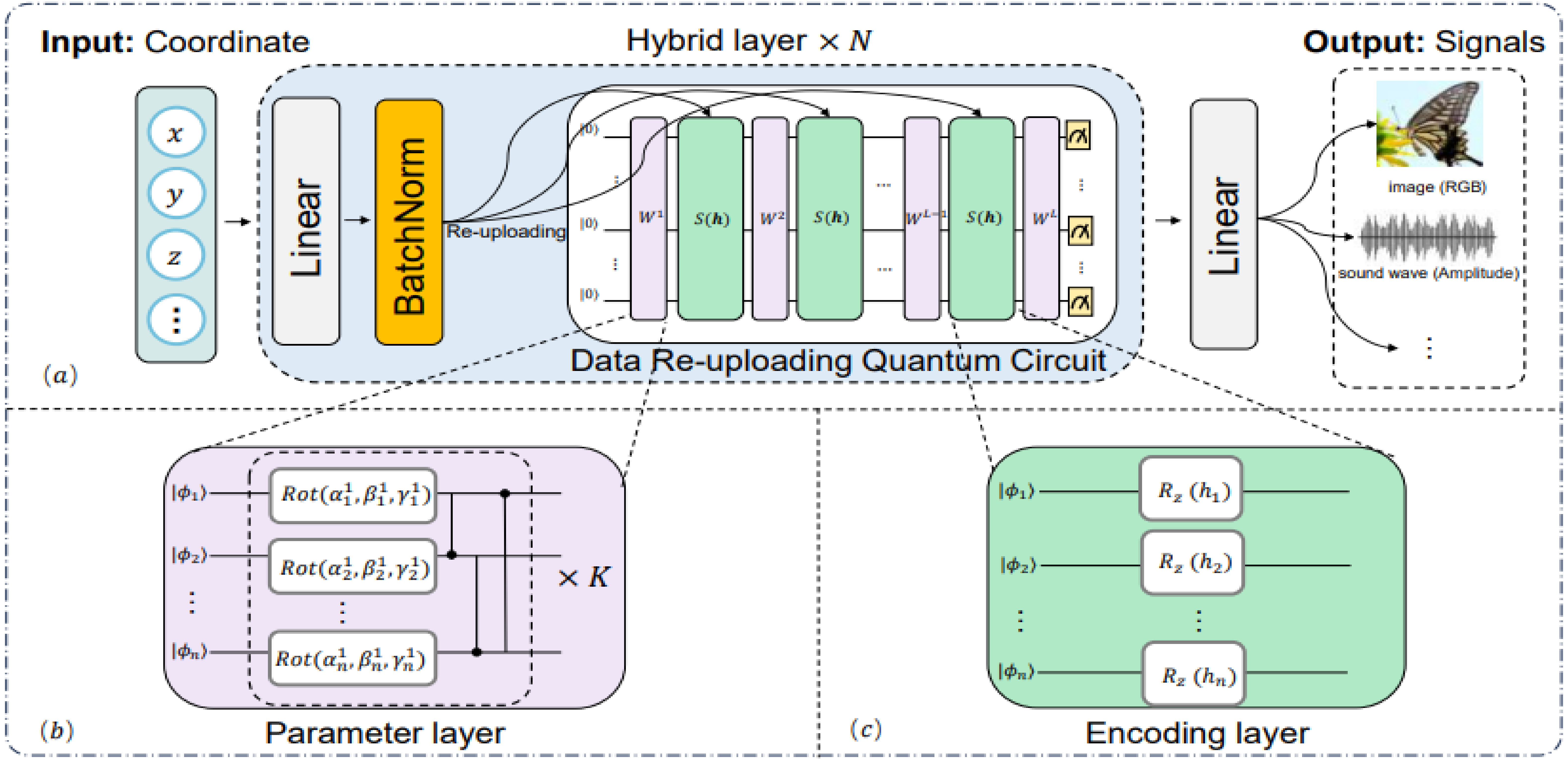



*Figure 3.* Architecture of QIREN. (a) presents the overall architecture of QIREN. (b) and (c) respectively illustrate the implementation details of the parameter layer and the encoding layer.

# Claim 1 - Quantum Advantage in Fourier Representation

- Classical INRs (like ReLU MLPs) struggle with high-frequency signals due to spectral bias.
- QIREN, using quantum circuits, has an **exponential advantage** in representing Fourier series.
- **Why?** The quantum data re-uploading mechanism enables a **denser frequency spectrum** with fewer parameters.
- This means QIREN captures fine details **more efficiently** than classical models.

# Claim 2 - Linear Layer Expands and Adjusts Frequency Spectrum

- QIREN uses a **pre-quantum linear layer** to optimize how frequencies are encoded.
- **Effect:**
  - Expands the frequency spectrum range
  - Reduces redundancy in learned frequencies
  - Helps the quantum circuit **better match real-world signals**
- This leads to **faster learning and better generalization**.

# Claim 3 - BatchNorm for Stability & Convergence

- Quantum circuits replace activation functions in QIREN.
- But training quantum models can be unstable!
- **Solution:** A **BatchNorm layer** before quantum processing:
  - Prevents vanishing/exploding gradients
  - Ensures **faster convergence** during training
  - Improves stability, especially in deep quantum networks
- Result: **QIREN trains reliably without complex hyperparameter tuning.**

# Datasets and Evaluation

| Task | Dataset | Details | Evaluation Metric |
|---|---|---|---|
| **Sound Representation** | Bach's Cello Suite | - 1000 equally spaced points<br>- Amplitudes normalized<br>- Timestamps in $[-1,1]$ | MSE |
| **Image Representation** | Astronaut, Camera, Coffee | - Images cropped & down-sampled to 32×32 pixels<br>- 1024 pixels per dataset | MSE |

**Models Used as Baselines:**

- **ReLU-based MLP**
- **Tanh-based MLP**
- **ReLU-based MLP with RFF** (Mildenhall et al., 2021)
- **SIREN** (Sitzmann et al., 2020)

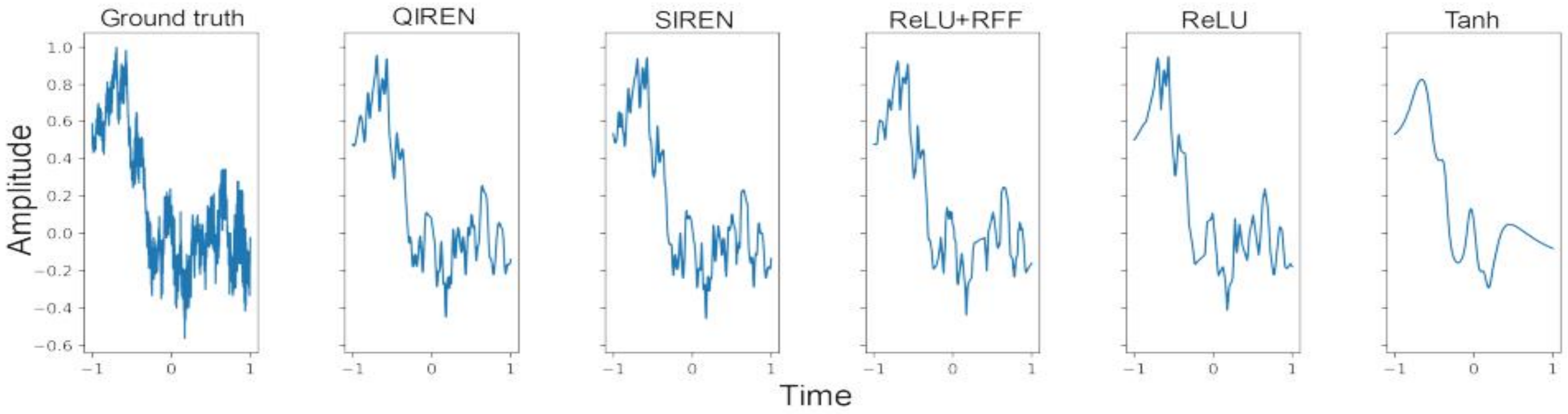



*Figure 4.* Results of sound representation. The Y-axis represents the normalized amplitude of the sound wave. The X-axis represents the time.


*Figure 5.* Results of image representation.

# RESULTS

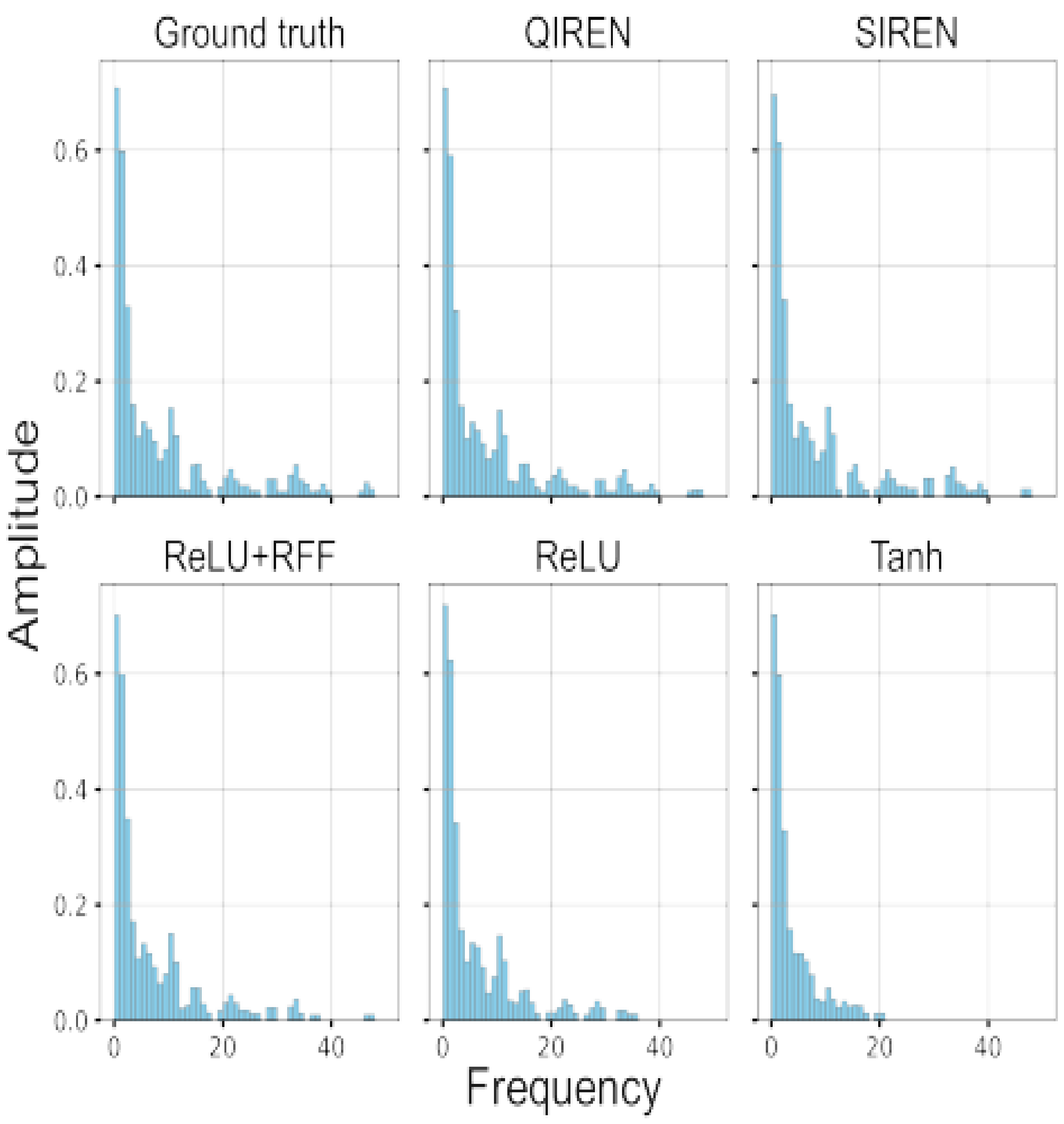



*Figure 6.* The frequency spectrum of the output of models on the sound representation task.

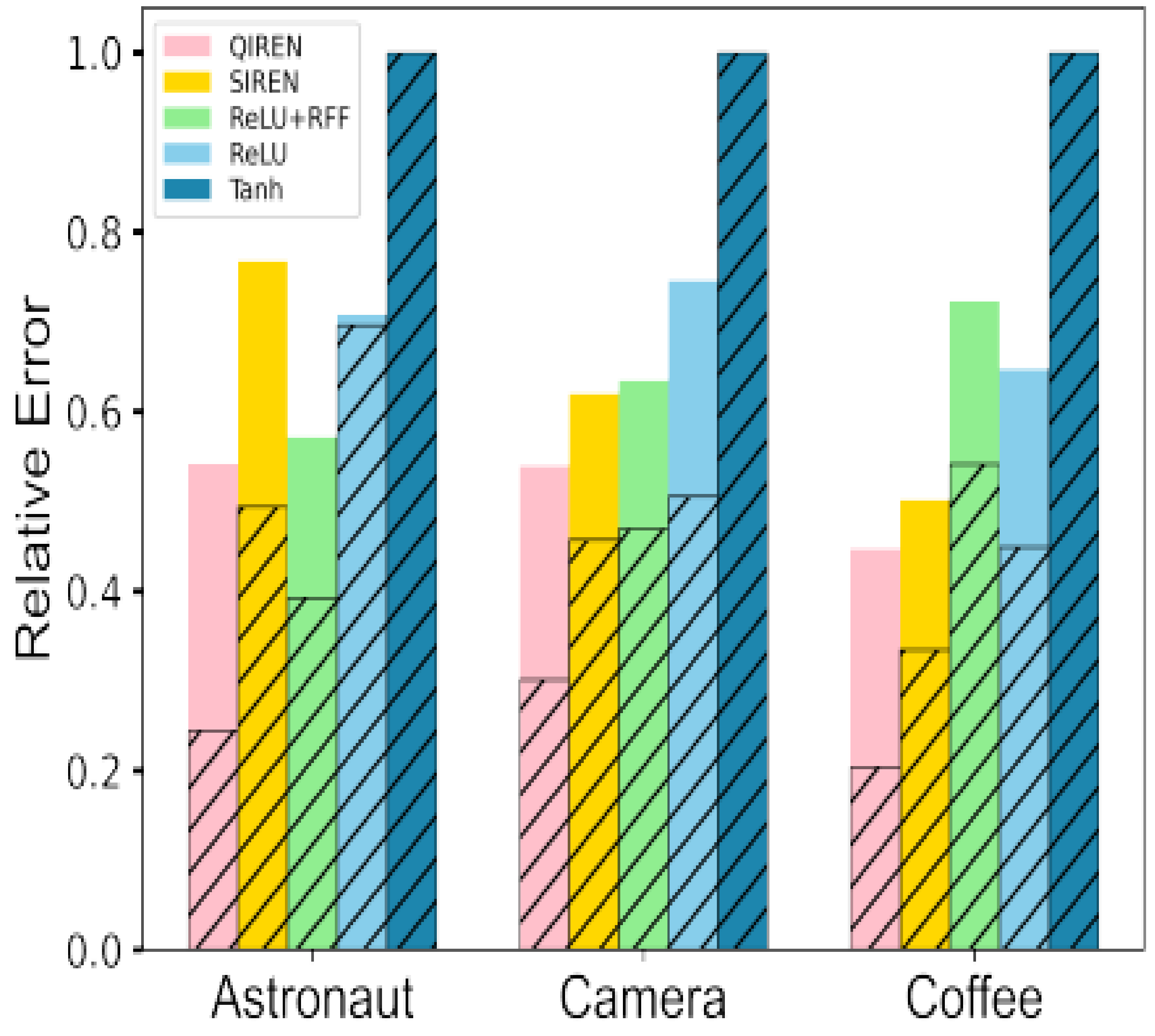



*Figure 7.* The relative error of each model compared to the Tanh-based MLP. The shaded area represents the low-frequency error, while the non-shaded area represents the high-frequency error.

- **Sound Representation**: QIREN saves **35.1% memory**, matches SIREN in performance, and excels in high-frequency representation.
- **Image Representation**: QIREN reduces error by **34.8%** with the fewest parameters.
- **Frequency Analysis**: QIREN achieves the lowest errors for high- and low-frequency components.

# Image Superresolution

| Dataset | Details | Ground Truth | Evaluation Metric |
|---|---|---|---|
| Astronaut, Camera, Coffee | Input: 32×32-pixel images- Partitioned into a 64×64 grid | 64×64 pixel images | MSE |

Additional models used: Bilinear interpolation, Nearest-neighbor interpolation.

# Results of Image SuperResolution

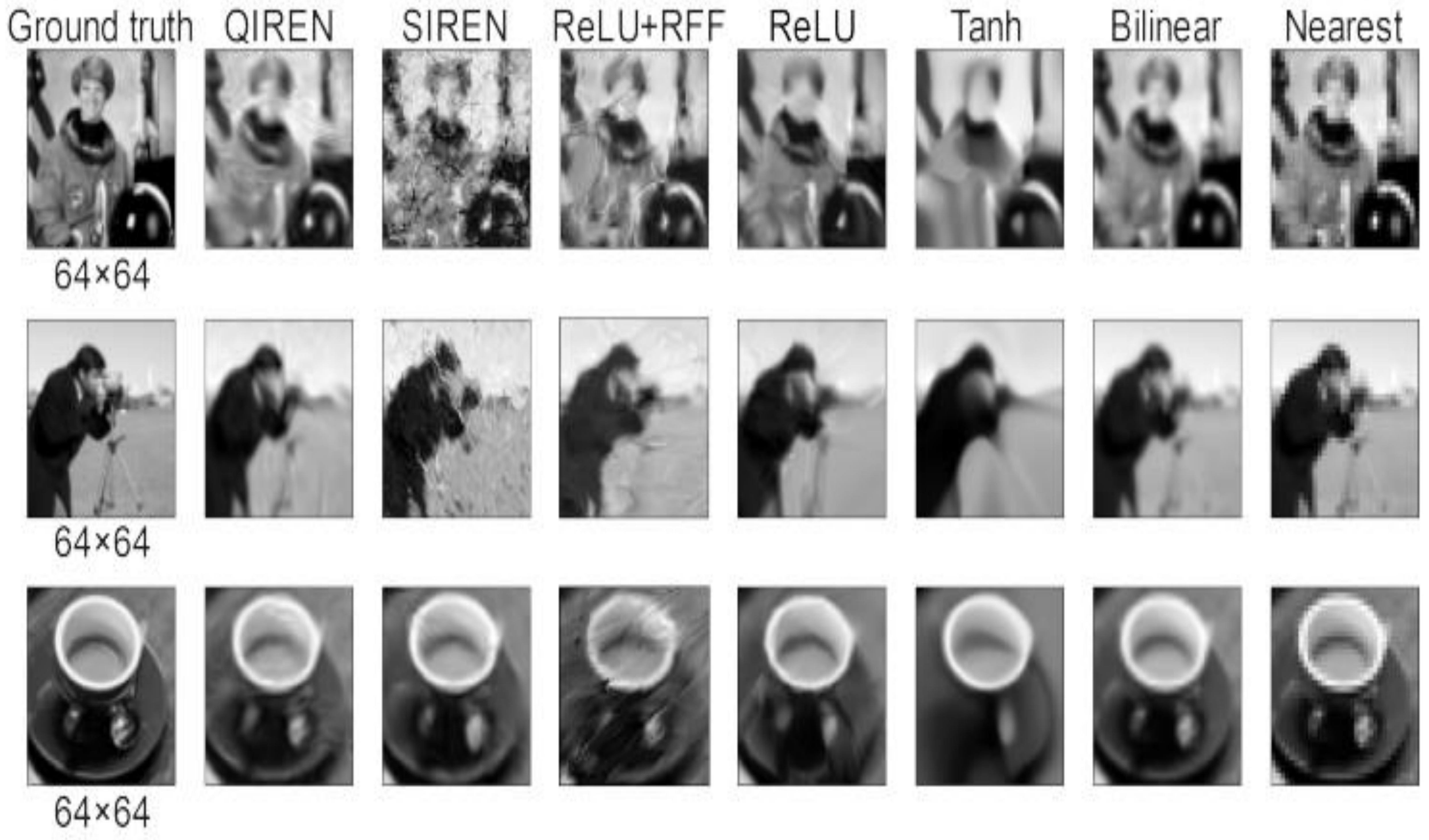



*Figure 8.* Results of image superresolution.

**Results:**

- QIREN reduces error by **24.0%**, outperforming interpolation methods.
- Classical Implicit Representation Networks perform worse than interpolation due to parameter limitations, but QIREN's superior signal representation aligns closely with the target function.

# Image Generation

**Datasets and Evaluation:**

Baseline models: Tanh-based MLP, ReLU-based MLP, ReLU+RFF, and SIREN.

| Dataset | Details | Resolution | Evaluation Metric |
|---|---|---|---|
| FFHQ | 70k human face images | 32×32 | Frechet Inception Distance (FID) |
| CelebA-HQ | 30k high-quality celebrity face images | 32×32 | FID (50k images for statistics) |

# Image Generation

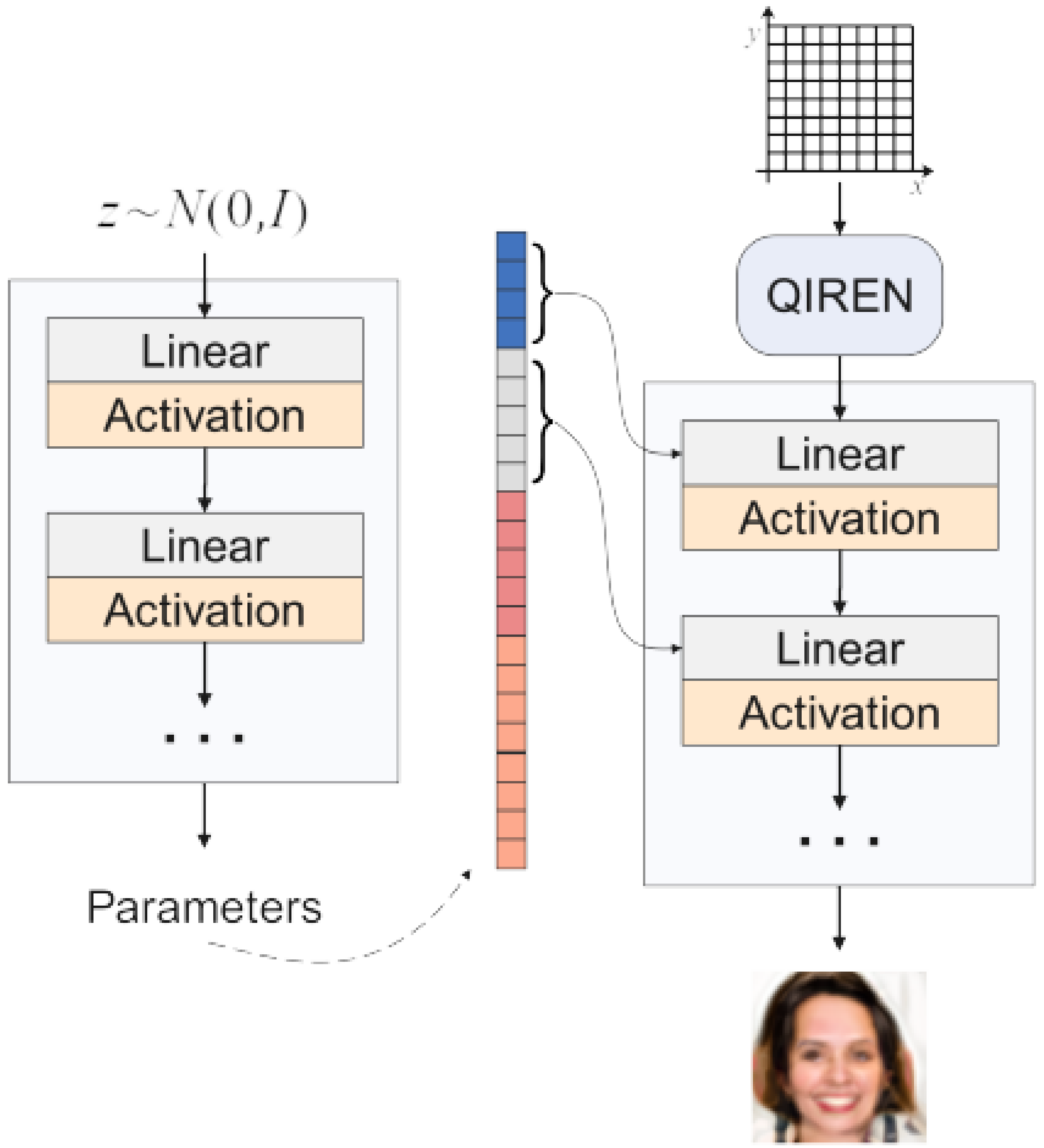



*Figure A1.* QIREN-based generator.

*Figure A2.* Results of image generation.

# Results of Image Generation

| Method | FFHQ | CelebA-HQ | #params |
|---|---|---|---|
| Tanh | 26.98 | 25.17 | 1.16M |
| ReLU | 84.94 | 110.81 | 1.16M |
| ReLU+RFF* | 15.01 | 13.91 | 1.14M |
| SIREN* | 22.31 | 20.97 | 1.16M |
| **QIREN (ours)** | **11.53** | **11.78** | **1.13M** |

*Table 2.* FID scores of different models on FFHQ and CelebA-HQ datasets.

# Conclusion and Future Work

- **QIREN** is derived as a **quantum generalization of FNNs**, with exponential advantages in representing Fourier series over classical FNNs.
- Experiments on **signal representation**, **image superresolution**, and **image generation** confirm QIREN's superior performance with fewer parameters compared to SOTA models.
- **Future Potential**: INRs, including QIREN, can be extended to applications like 3D object representation, time series forecasting, and solving differential equations.
- This work highlights a practical scenario for **quantum machine learning**, paving the way for further research and innovation in this field.

# References

Zhao, J., Qiao, W., Zhang, P., & Gao, H. (2024). Quantum implicit neural representations. *arXiv preprint arXiv:2406.03873*.

# THANK YOU

## FID Formula

$$\mathrm{FID} = ||\mu_r - \mu_g||^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

Where:

- $\mu_r, \Sigma_r$ → Mean and covariance of real images
- $\mu_g, \Sigma_g$ → Mean and covariance of generated images
- $||\mu_r - \mu_g||^2$ → Squared difference between means
- $\mathrm{Tr}(\cdot)$ → Trace of a matrix

### Typical FID Scores

- **FID < 10** → High-quality, nearly indistinguishable from real images
- **10 ≤ FID < 50** → Good but noticeable differences
- **FID > 50** → Poor image quality

## A. Preliminaries of Quantum Circuit

In quantum computing, information is often carried by qubits over Hilbert space. A pure quantum state consists of one or more qubits and is usually represented by Dirac's notation, which denotes a unit vector $\mathbf{v}$ as a ket $|v\rangle$ and its conjugate transpose $\mathbf{v}^\dagger$ as a bra $\langle v|$. The inner product between $|v\rangle$ and $|u\rangle$ is denoted as $\langle u|v\rangle$, and the outer product is $|u\rangle\langle v|$. The evolution of a quantum state $|v\rangle$ is accomplished by sequentially applying quantum gates on it, i.e. $|v'\rangle = U_K...U_2U_1|v\rangle$, where $U_k$ is the unitary matrix representing the quantum gate and $|v'\rangle$ is the quantum state after evolution. Common single-qubit gates are as follows:

$$\begin{aligned}
H &:= \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, I := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\
X &:= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, R_X(\theta) := \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \\
Y &:= \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, R_Y(\theta) := \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}, \\
Z &:= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, R_Z(\theta) := \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix},
\end{aligned} \tag{A1}$$

where $H$ denotes the Hadamard gate, $X, Y, Z$ denote the Pauli gates, $R_X(\theta), R_Y(\theta), R_Z(\theta)$ denote the rotation gates. A common multi-qubit gate is $CNOT$ gate:

$$CNOT := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{A2}$$

In a quantum circuit, the initial quantum state is generally $|0\rangle^{\otimes N}$, and after applying a sequence of quantum gates, the measurement will be used to convert quantum information into classical information. For instance, we can design quantum measurements to obtain the expectation $\langle v|O|v\rangle$ of the quantum state $|v\rangle$ about an observable $O$.