Quantum Error Mitigation via Autoencoder Neural Networks

Xiao-Dao Lin[1], Hsi-Ming Chang[2], Jhih-Shih You[3] and Hsiu-Chuan Hsu[1]

[1] Graduate Institute of Applied Physics, National Chengchi University

[2] Brightlight Vision, Santa Clara, CA

[3] Department of Physics, National Taiwan Normal University

Content

- Introduction
- Methodology
- Results & Discussion
- Conclusion & Future work

Introduction

Motivation, Previous work & research

Background – Motivation & Previous work

- Due to the challenging control of qubits in different techniques, there are some environmental perturbations that cause the errors in measuring the qubits (readout errors) .

Kevin J. Sung, Quantum Algorithm Engineer, IBM, "Execution on Noisy Quantum Hardware," *Road to Quantum Utility Workshop 2025: Using Quantum Devices with More Than 100 Qubits*, January 22, 2025. Image from IBM Quantum.

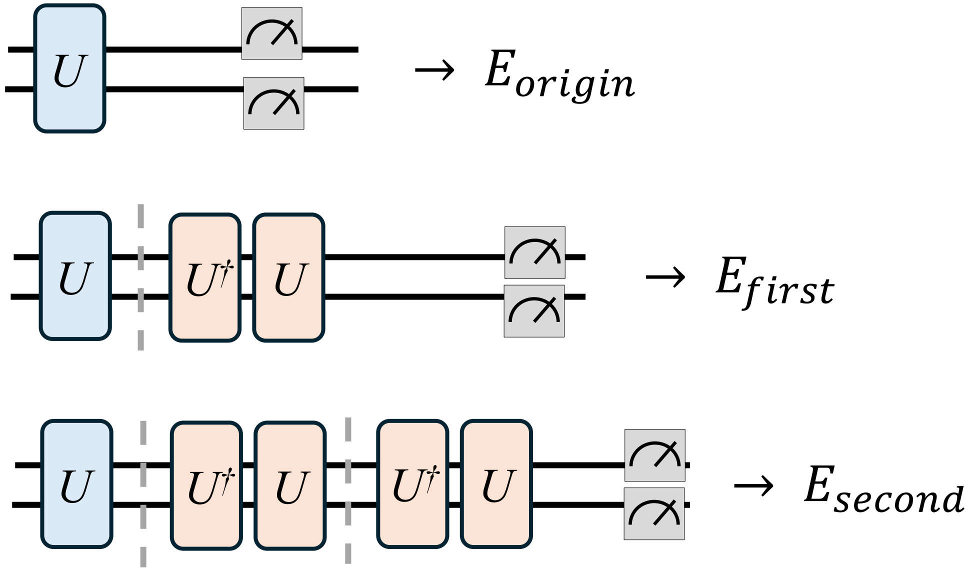
- A promising approach is quantum error mitigation which focuses on reducing errors. Common methods include zero-noise extrapolation(ZNE), probabilistic error cancellation, measurement error mitigation and learning-based methods [2].

[1] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, “Hybrid quantumclassical algorithms and quantum error mitigation,” Journal of the Physical Society of Japan, vol. 90, no. 3, p. 032001, 2021. [Online]. Available: https://doi.org/10.7566/JPSJ.90.032001
[2] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O’Brien, “Quantum error mitigation,” Rev. Mod. Phys., vol. 95, p. 045005, Dec 2023. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.95.045005

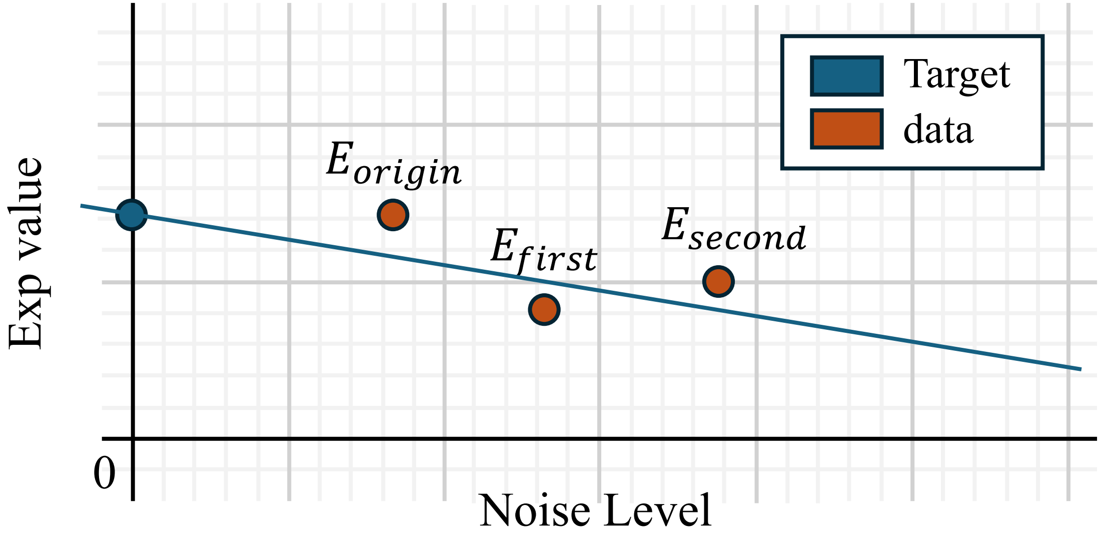
Background – zero-noise extrapolation

- Two steps of ZNE:

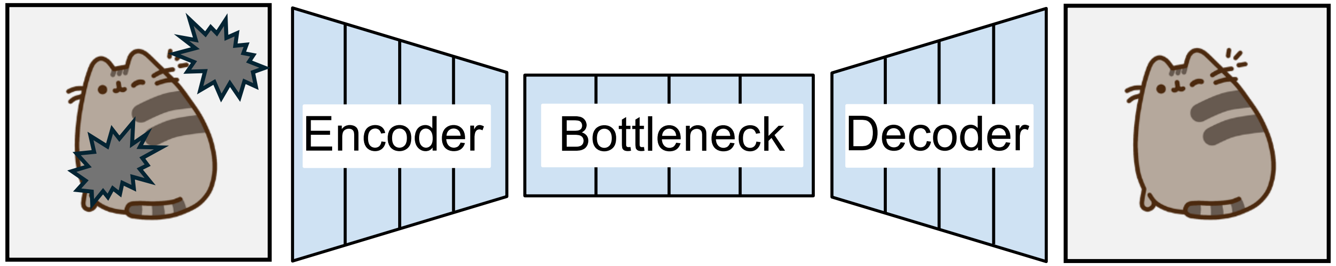
1. Intentionally scale noise



2. Extrapolation to the noiseless limit



Background – Autoencoder



- Vincent et al. [3], [4] introduced denoising autoencoders, demonstrating their capability to extract and compose robust features by reconstructing original signals from corrupted inputs.

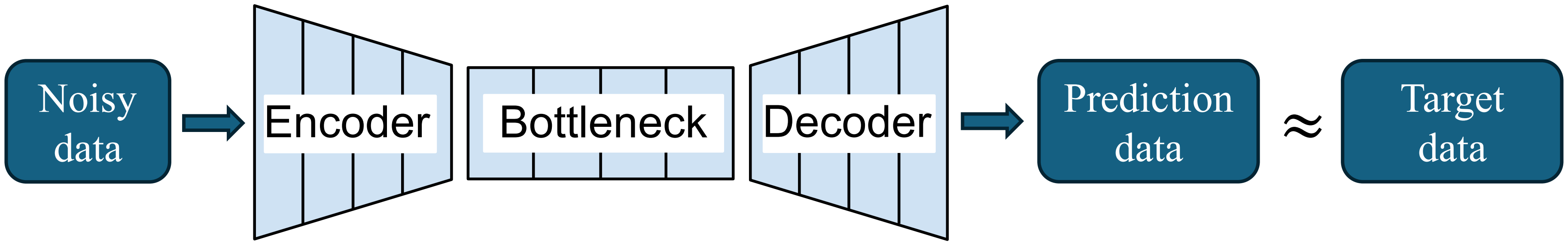
- Zhang [5] shows the effectiveness of convolutional autoencoders in image denoising tasks, demonstrating superior spatial feature learning compared to traditional fully connected architectures.

[3] P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol, "Extracting and composing robust features with denoising autoencoders," Proceedings of the 25th International Conference on Machine Learning, pp. 1096– 1103, 2008.
[4] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," Journal of Machine Learning Research, vol. 11, pp. 3371–3408, 2010.
[5] Y. Zhang, "A better autoencoder for image: Convolutional autoencoder," in Proceedings of the International Conference on Image Processing, 2018.

This study

- In this paper, we propose to use convolutional denoising Autoencoder to mitigate the readout errors in measurement probabilities of quantum circuits.

Noisy data	Targets data
$\begin{bmatrix} 0.3 \\ 0.35 \\ 0.2 \\ 0.15 \end{bmatrix}$	$\begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$



Methodology

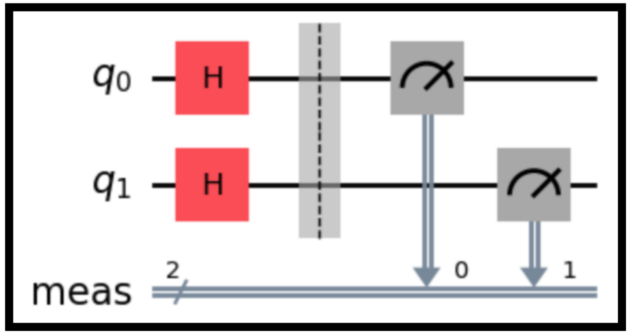
The methods.

Quantum Measurement Probability

- To control the wave function $|\psi\rangle$, we design a circuit for specific number of qubits by adding operators(gates).

- The measurement probability is given by $|\langle i|\psi\rangle|^2$.
  ($|i\rangle$ stands for one of the basis vectors).

- If the wave function is $|\psi\rangle = H \otimes H|00\rangle = \frac{1}{\sqrt{4}}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$:



One shot

$00, 01, 10, 11$

00 **(Readout)**

Measurement in 1k shots

(One experiment)

$$\begin{bmatrix} 0.24 \\ 0.253 \\ 0.278 \\ 0.229 \end{bmatrix} \approx \begin{bmatrix} |\langle 00|\psi\rangle|^2 \\ |\langle 01|\psi\rangle|^2 \\ |\langle 10|\psi\rangle|^2 \\ |\langle 11|\psi\rangle|^2 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

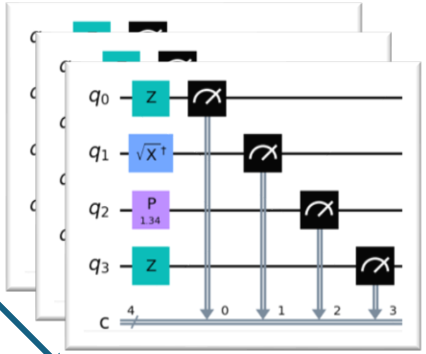
(Measurement Probability)

Datasets Generation – Qiskit SDK

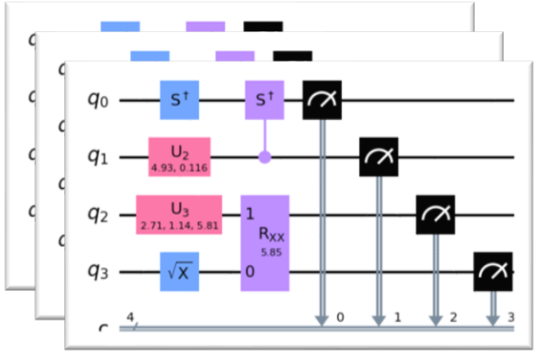
- Circuits : 700 four qubits **Random** circuits for each **depth** from 1 to 18.
- Measurement probabilities for each circuits to get the target and noisy data:

Datasets	Experiment for each circuit	Shot per experiment	Backend (target & input)	Size of row data
1E1Ks	1	1K	Aer & FakeLima	700*18 = 12.6k
100E1Ks	100(increase)	1K	Aer & FakeLima	1.26M
100E10Ks	100	10K(increase)	Statevector & FakeLima	1.26M

Depth1

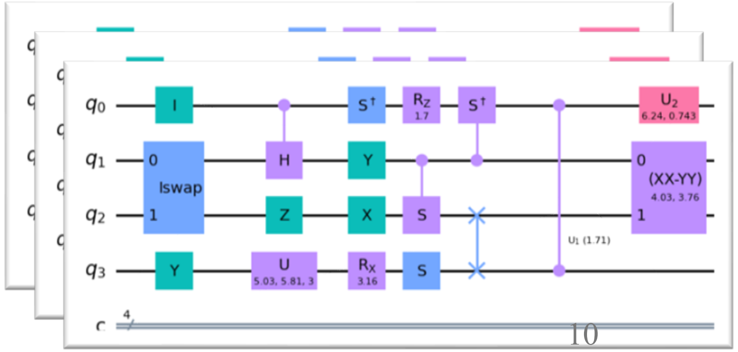


Depth2



. . .

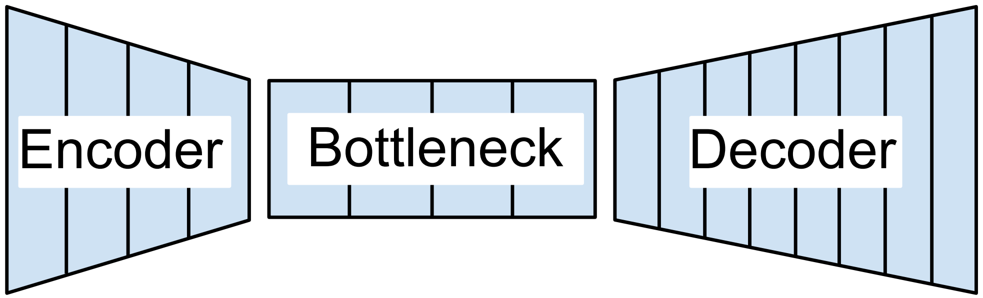
Depth 18



700 independent circuits

Models – Autoencoder

- In our work, using 1D Convolutional layers instead of 2D layers contains the same performance without extra data transformation.
- Below is the schematic plot to illustrate the model we use.



Training Process

- During training:

Type	Method
Cost function	Mean Squared Error
Optimizer	Adam
Learning rate	0.0005
Training Epoch	500 epochs
Batch size	10k
Saved Checkpoint	At Minimal validation loss

- Software & Hardware:

Type	Method
Code implementation	Tensorflow
CPU(for data generation)	Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz (6 cores, 12 threads)
GPU(for training model)	NVIDIA TITAN RTX GPUs (24GB VRAM each)

Model evaluation

- Loss calculation method : Two metrics about Mean Absolute error(MAE).

$$Dataset\ error\text{: } L_{MAE} = \frac{1}{N}\sum_{j=1}^{N} |P_{targ,j} - P_{noise,j}|,$$

$$Prediction\ error\text{: } L_{MAE} = \frac{1}{N}\sum_{j=1}^{N} |P_{targ,j} - \hat{P}_j|$$

*N is the number of basis vectors ; $P_{targ,j}$ & $P_{noisy,j}$ is the target and noisy measurement probability for j-th basis vector ; $\hat{P}_j$ is the prediction measurement probability for j-th basis vector.

Model evaluation

- We use four common quantum circuits and algorithms as the evaluation datasets.

(a) Trivial Paramagnet

(b) A single depth of Random unitary gates drawn from Haar measure (Haar single)

(c) Grover's search algorithm with the target state $|1111\rangle$

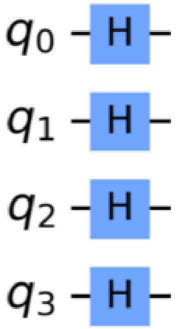
(d) Quantum Fourier Transformation (QFT)

Model evaluation

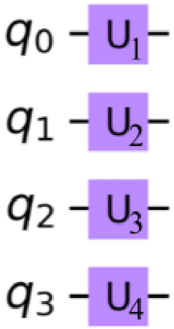
U_0 is the u-gate with (ϑ,ϕ,λ) in $(\frac{\pi}{2}, 0, \pi)$.
$U_{1\text{-}4}$ is the u-gate with (ϑ,ϕ,λ) sample from Haar measure.

$$U(\theta,\phi,\lambda) = \begin{bmatrix} \cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2} \end{bmatrix}$$

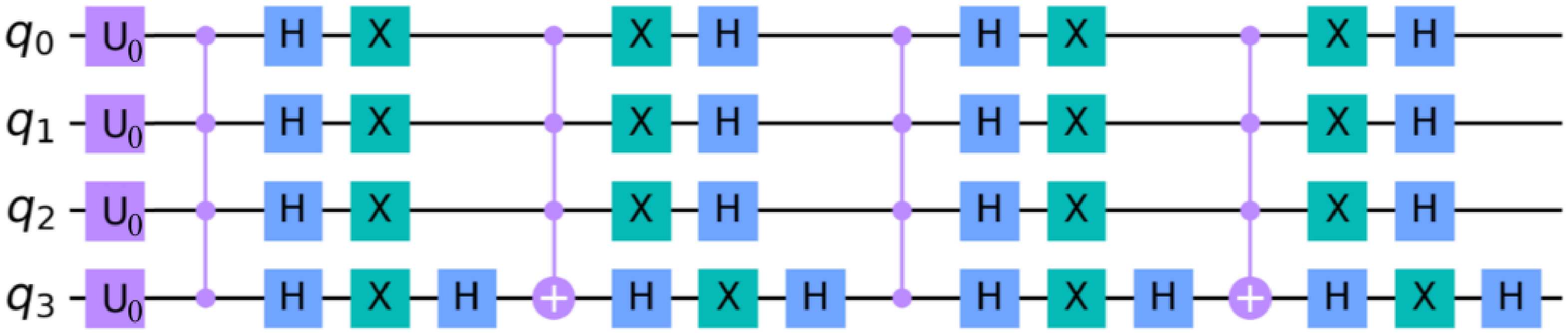
(a)Trivial Paramagnet



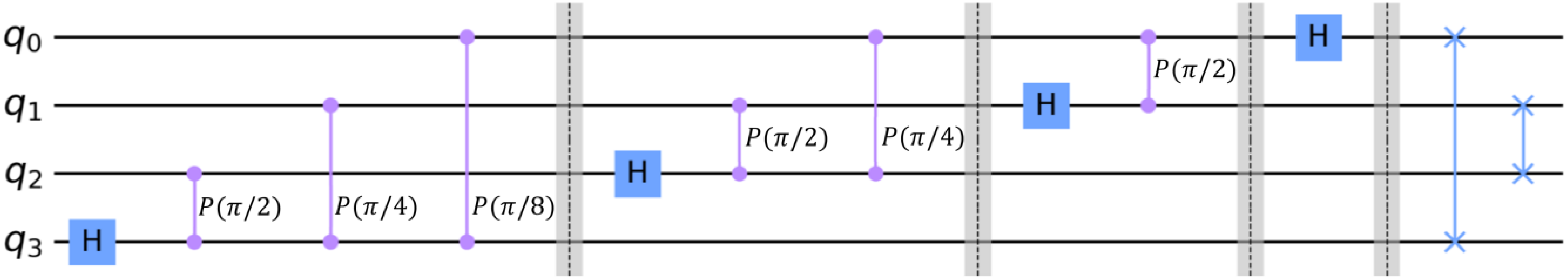
(b) Haar Single



(c) Grover's Search



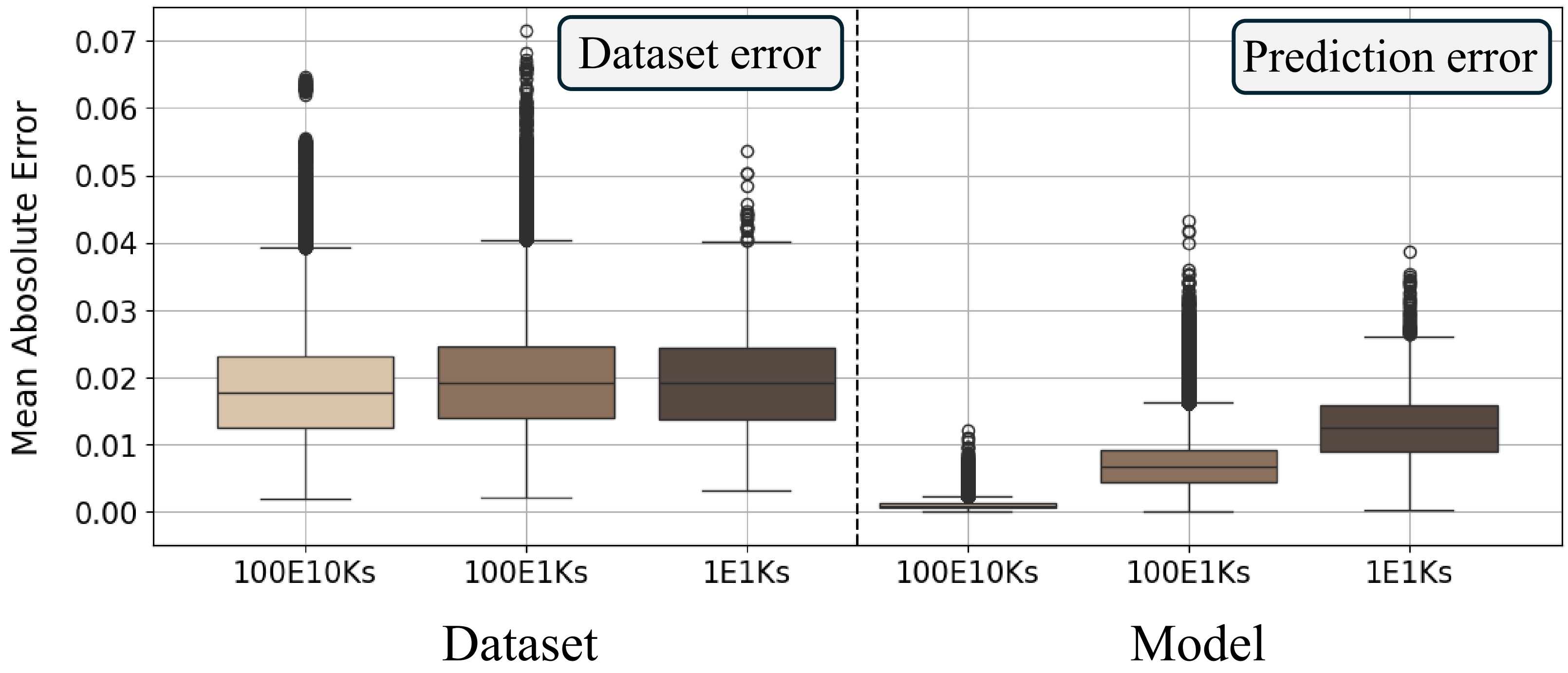
(d) QFT



Results

Discuss about the results.

Datasets noise level & Prediction error

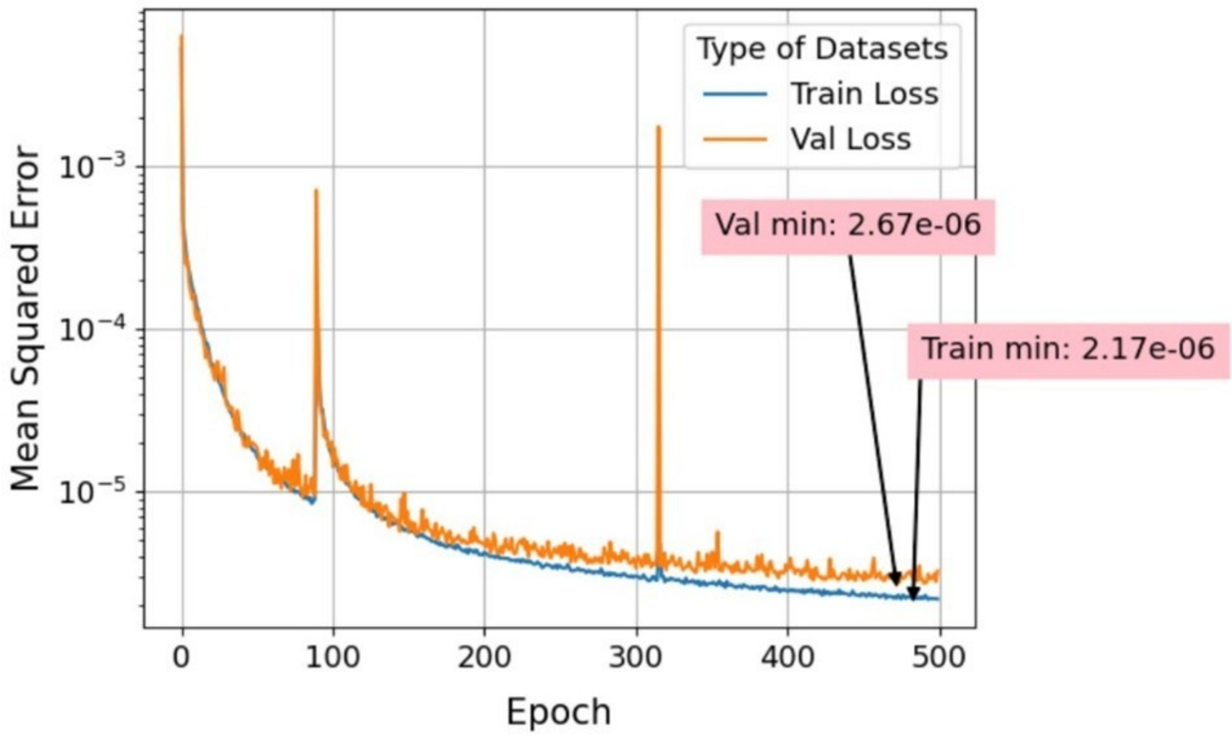


Dataset naming rule:
E: experiment
K: shot number

- The noise level for the three test datasets is almost the same.
- More training datasets improve the prediction. With more shots per experiment, the prediction improves further.

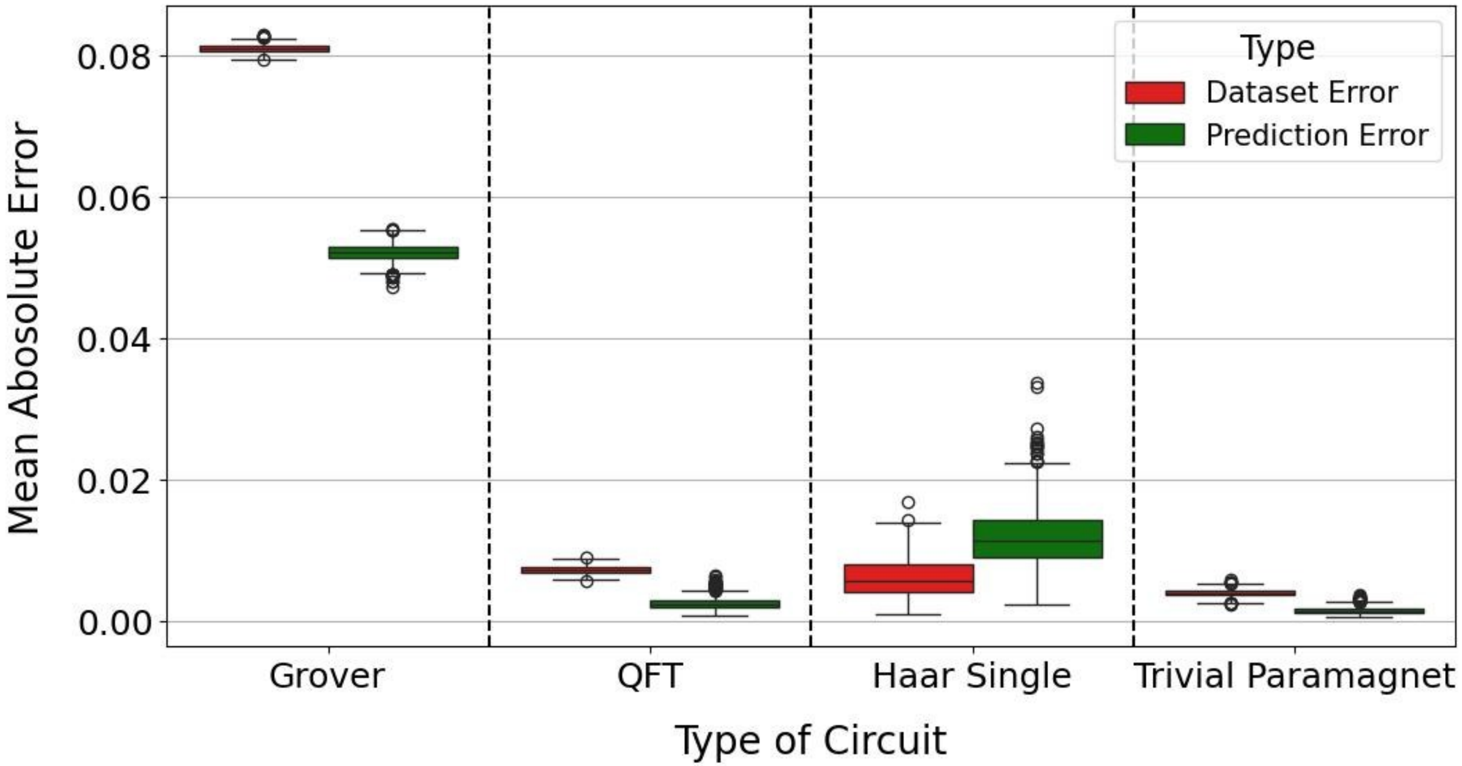
Take model trained by **100E10Ks** for evaluating part.

Loss curve



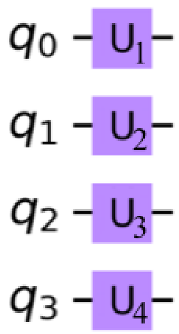
- No overfitting.
- The validation minimum reaches 10^{-6}.

Evaluating with unknown circuits

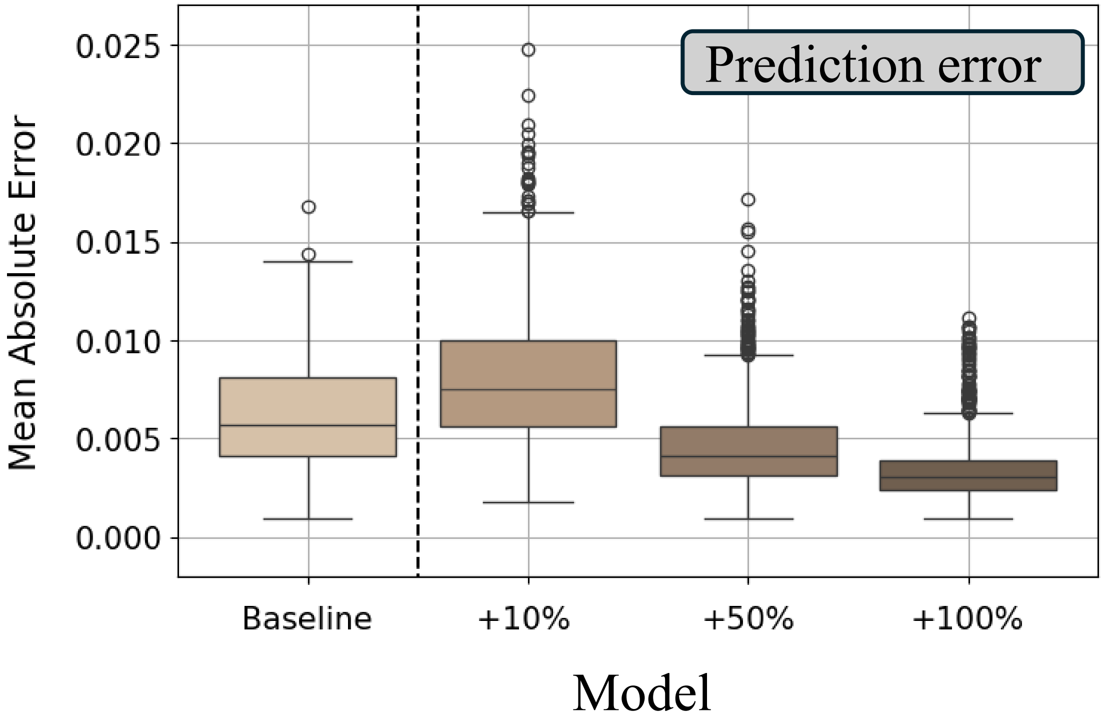


- For Grover's Search, Quantum Fourier Transform (QFT), and trivial paramagnet, our model denoises the error at different levels.
- However, for Haar single, our model increases the error instead.

Bias datasets for Haar single



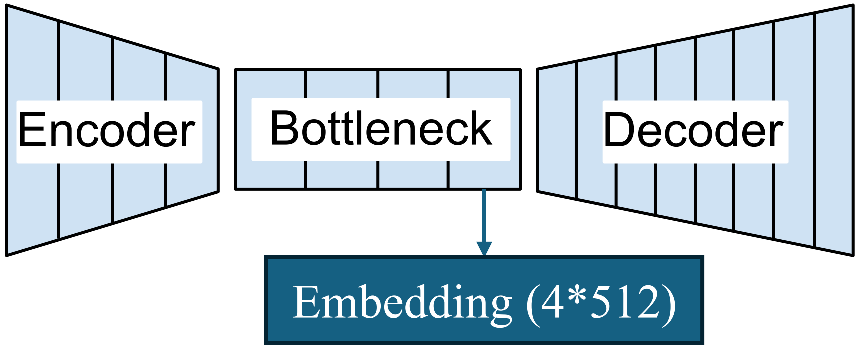
- The datasets lack information about unentangled circuits.
- We generate additional Haar single datasets to bias and add into the 10E10Ks datasets by different ratios.

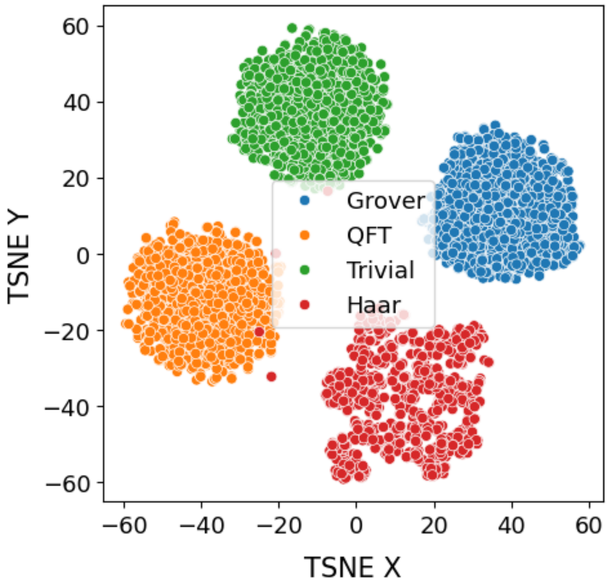


T-SNE Visualization

- For four algorithm datasets inputs, we take the output from the last bottleneck layer and reduce it to two dimensions using t-SNE.

- A clear classification into four distinct groups for the four different algorithms.

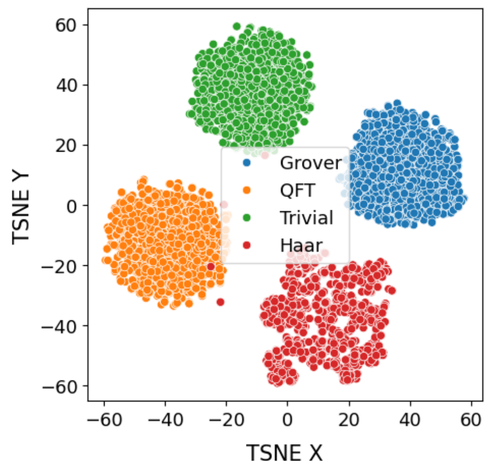




Conclusion

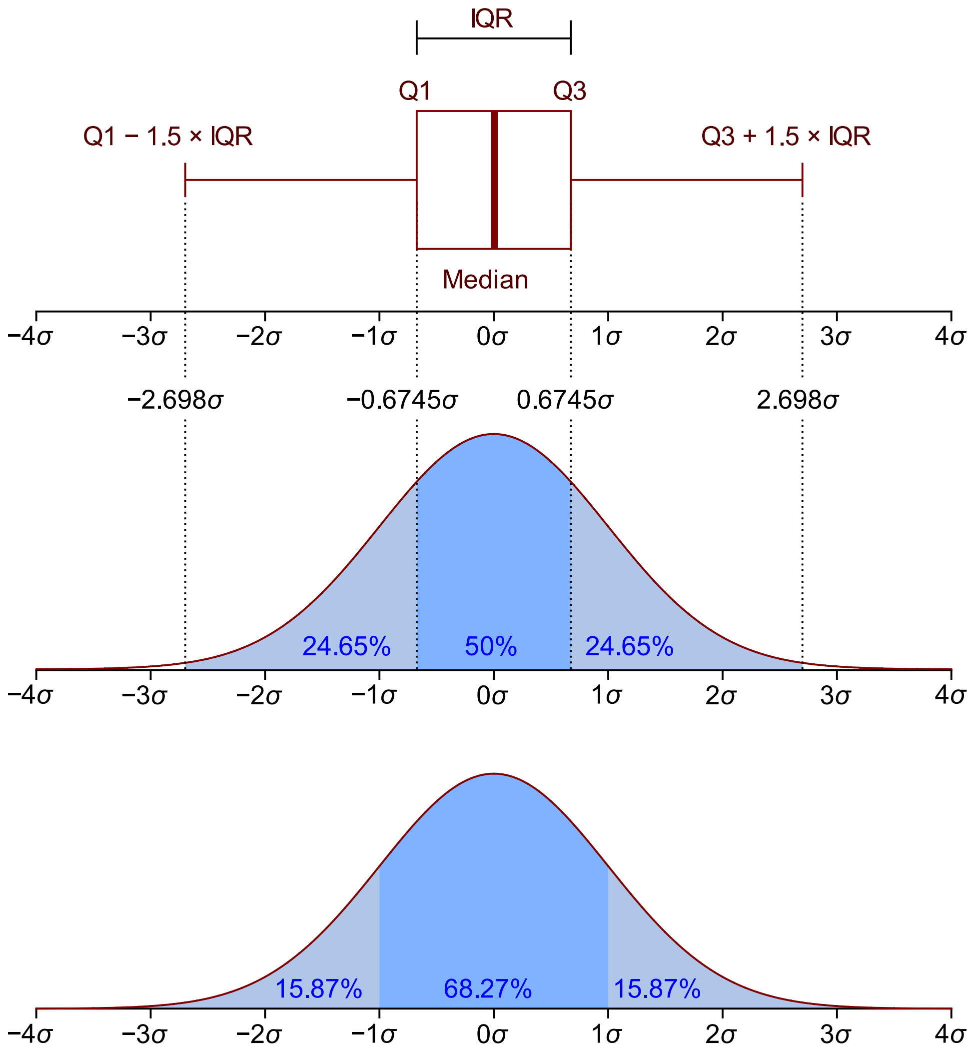
The conclusions.

Summary



- Take-home message
  - The statistical error in both target and noisy data sharply influences the convergence of the model.
  - The autoencoder can reduce the noise in datasets from the noise model.
  - For unseen circuits, the model mitigates the noise at different levels.
  - The encoder successfully learns the features of the four common circuits.
- Future work
  - The noise model cannot reflect the noise level in real quantum machines, so we expect the model need to be fine-tuned with the real quantum machine.
  - Accommodating different lengths of measurement probability inputs using a sequence-to-sequence model.

Appendix A - Boxplot



From Jhguch at en.wikipedia, CC BY-SA 2.5,
https://commons.wikimedia.org/w/index.php?curid=14524285