FORDHAM UNIVERSITY
THE JESUIT UNIVERSITY OF NEW YORK

ATP: Adaptive Threshold Pruning for Efficient Data Encoding in Quantum Neural Networks

Conference on Computer Vision and Pattern Recognition (CVPR), 2025.

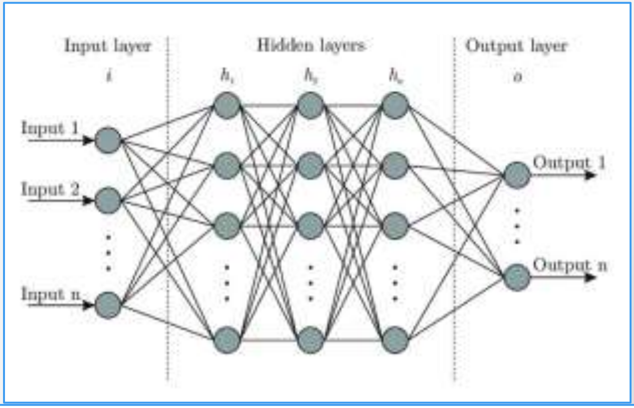
Mohamed *Afane*[1], Gabrielle Ebbrecht[1], Ying Wang[2], Juntao Chen[1], and Junaid Farooq[3]

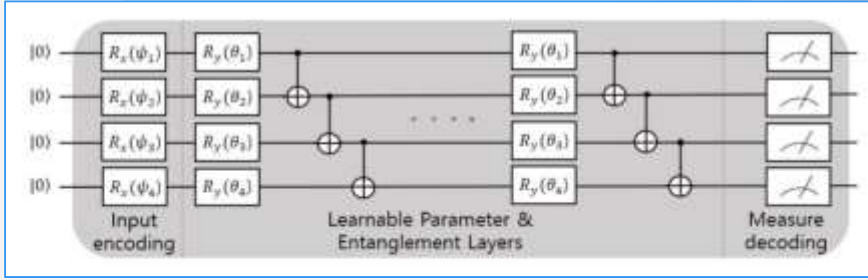
[1]Fordham University [2]Stevens Institute of Technology [3]University of Michigan-Dearborn

Background

Quantum Neural Networks (QNNs) vs. Classical Neural Networks

- QNNs use quantum circuits to process information, unlike classical NNs that rely on matrix multiplications..
- QNNs can exploit entanglement to learn representations differently from classical networks.
- Potential advantages: higher expressive power & improved efficiency for certain tasks.





Superposition

Qubits exist in linear combinations of 0 and 1 states, represented as:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle,$$

Where α and β are complex numbers that satisfy the constraint $|\alpha|^2 + |\beta|^2 = 1$

This property is ***superposition***

- Allows for concurrent calculations
- Algorithms are inherently parallel

Entanglement

- The state of one qubit directly influences the state of another
- Not constrained by distance
- Enables intricate encoding, but challenging to preserve

An entangled state can be expressed as:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Entanglement Entropy (EE)

- Quantifies correlation between different parts of the system
- Reflects a model's capacity to capture intricate relationships

LOW

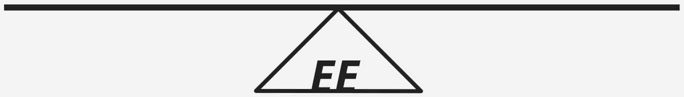
Efficient resource usage

Reduced complexity

HIGH

Qubit interdependence

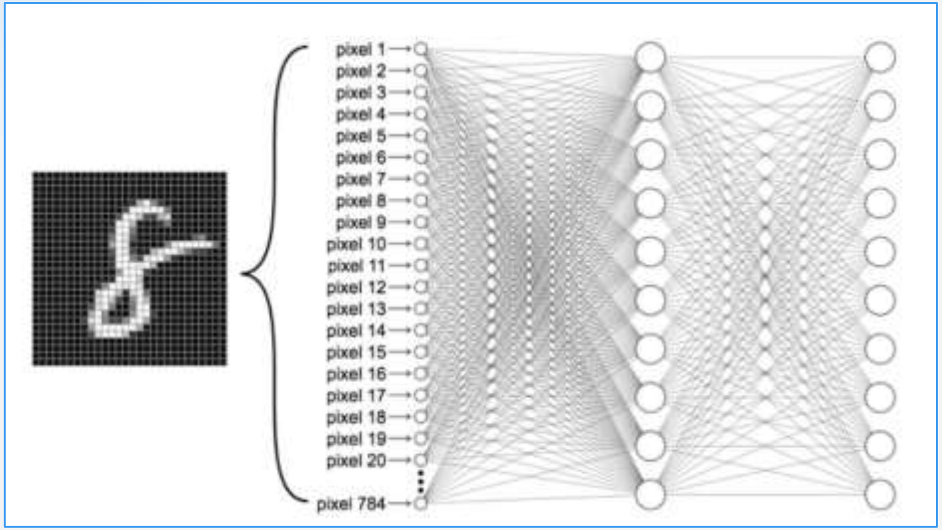
Potential overfitting

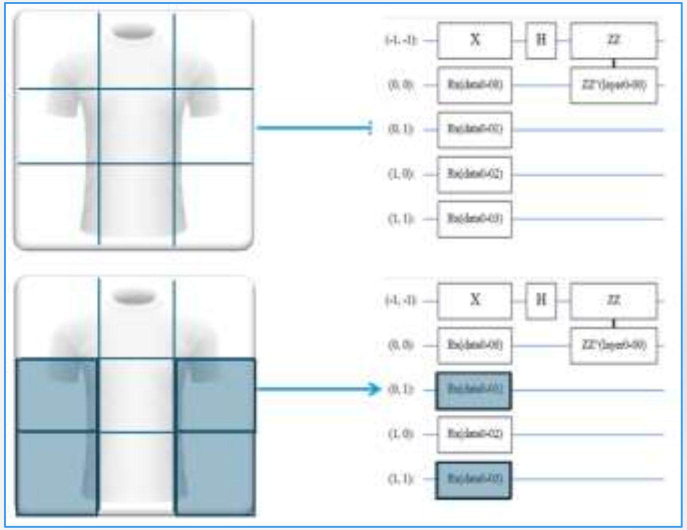


Need some balance!

Introduction

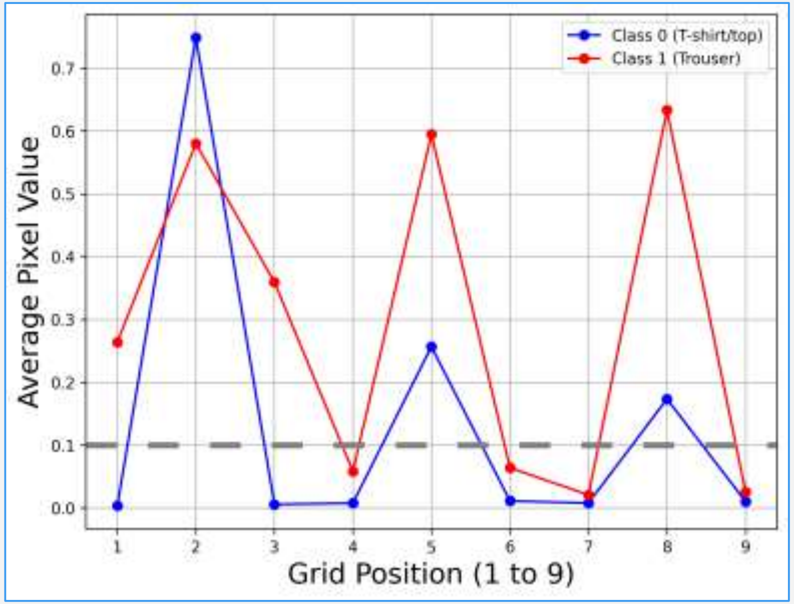
How are images processed by computers?





Introduction

How are images processed by computers?



Introduction

In this work, ATP is applied *before* encoding data into quantum states.

Angle Encoding

Each pixel value $x_{i,j}$ of an image is mapped onto an angle $\theta_{i,j}$ for the RX gate:

$$|x\rangle = \bigotimes_{i=1}^{N} \cos(x_i)\,|0\rangle + \sin(x_i)\,|1\rangle$$

→ Straightforwardly represents pixel intensities, but can introduce redundancy for higher-dimensional data

Amplitude Encoding

Encodes normalized data values as amplitudes of a quantum state

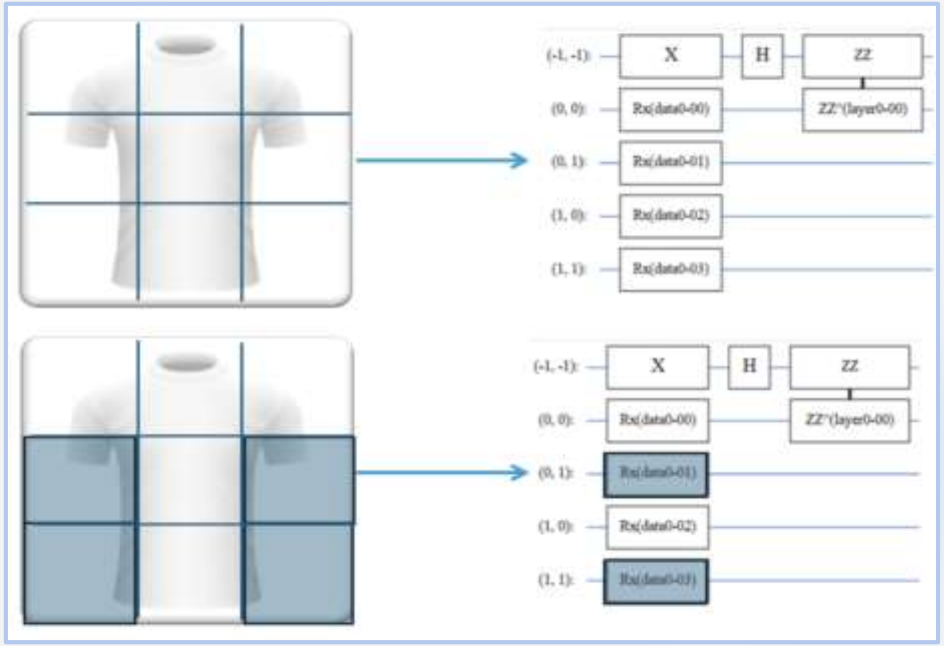
$$|\psi\rangle = \sum_{i=0}^{2^n-1} x_i|i\rangle$$

→ Faces scalability challenges for large datasets

Adaptive Threshold Pruning

Optimize encoding process by adaptively pruning non-essential features

- **Reducing qubit usage and entanglement** in the circuit
- Set **dynamic thresholds** for feature pruning
- Selectively **remove low-variance data** to maintain model performance while lowering entanglement entropy

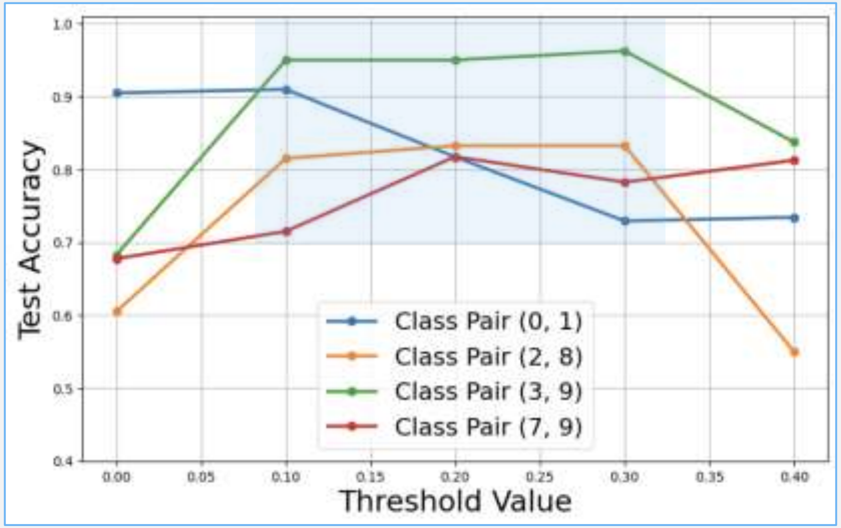


Data Pruning Effects on Accuracy

Moderate pruning is generally optimal

- Ranging from **0.1 to 0.3** for MNIST
- Eliminates non-essential features
- Preserves critical distinctions

Higher thresholds reduce accuracy by removing valuable details.



Test accuracy when distinguishing various class pairs from MNIST

$$x_\tau(i,j) = \begin{cases} 0, & \text{if } \bar{x}_0(i,j) < \tau \text{ and } \bar{x}_1(i,j) < \tau, \\ x(i,j), & \text{otherwise,} \end{cases}$$

EXPERIMENTS

Results for various encoding techniques

Setup

A 3-layer QNN model was applied on several datasets using various techniques:

Benchmark Datasets
MNIST
FashionMNIST
CIFAR
PneumoniaMNIST

Encoding Techniques
Angle
Amplitude

Preprocessing	
ATP	Refine data structure before encoding
PCA	
SQE	Focus on single-qubit encoding

Noise Conditions

- Introduced depolarizing noise at intensities of 3%, 5%, and 10% to evaluate robustness in realistic settings
- Allows for comparison of baseline performance and resilience to noise

Performance Results

Accuracy

Classes	Angle	Amplitude	ATP	PCA	SQE
MNIST					
(0,1)	96.0	95.5	**99.0**	**99.0**	88.0
(0,3)	89.0	88.5	**91.0**	88.0	86.0
(2,4)	85.0	84.0	**86.0**	84.5	82.0
(5,6)	86.0	85.5	**87.0**	85.0	83.5
(2,8)	81.0	79.5	83.0	**86.0**	78.5
Fashion MNIST					
(0,1)	88.5	88.0	**91.5**	88.5	86.0
(2,8)	**86.0**	84.5	**86.0**	**86.0**	83.0
(3,9)	**94.0**	87.0	**94.0**	93.0	91.0
(7,9)	82.0	78.0	**83.0**	79.0	77.0
CIFAR					
(0,1)	70.0	68.5	**74.2**	68.0	66.0
PneumoniaMNIST					
(0,1)	81.0	68.5	**87.0**	80.0	75.5

Entanglement Entropy

Classes	Angle	Amplitude	ATP	PCA	SQE
MNIST					
(0,1)	0.67	0.52	**0.39**	0.60	0.45
(0,3)	0.59	0.53	**0.34**	0.55	0.44
(2,4)	0.77	0.64	0.45	0.56	**0.41**
(5,6)	0.82	0.53	0.41	0.61	**0.39**
(2,8)	0.63	0.56	**0.34**	0.36	**0.34**
Fashion MNIST					
(0,1)	0.56	0.47	**0.35**	0.58	0.39
(2,8)	**0.52**	0.50	0.35	0.41	0.38
(3,9)	0.54	0.62	**0.38**	0.55	0.42
(7,9)	0.59	0.58	0.41	0.64	**0.43**
CIFAR					
(0,1)	0.63	0.51	0.46	0.65	**0.43**
PneumoniaMNIST					
(0,1)	0.88	0.79	**0.37**	0.59	0.42

Across the majority of datasets and class pairs, ATP achieves the **highest accuracy** and consistently **minimizes EE** → lowers complexity without compromising accuracy

Performance Results: *Depolarizing Noise*

Accuracy

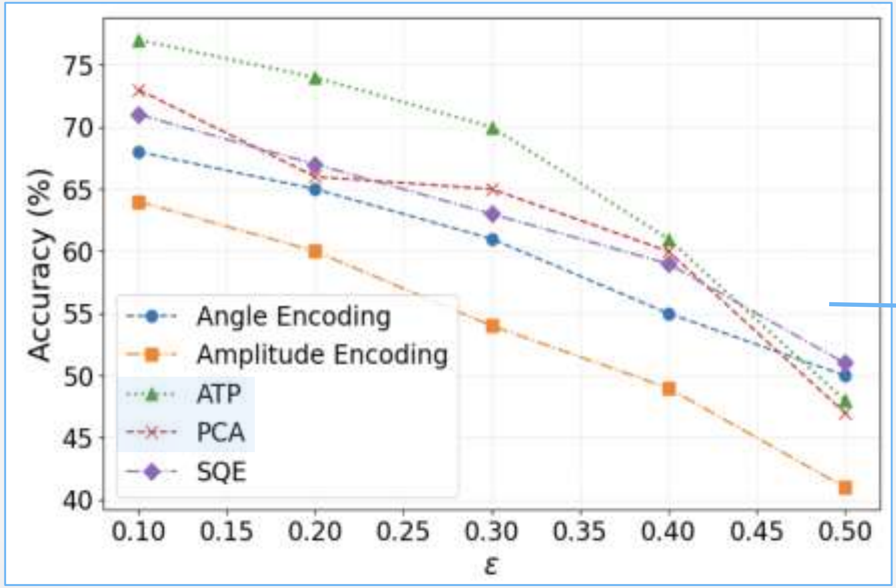
Classes	Angle	Amplitude	ATP	PCA	SQE
MNIST					
(0,1)	87.0	86.4	**89.0**	**89.0**	88.2
(0,3)	82.4	80.5	**83.0**	78.9	81.9
(2,4)	**80.2**	68.8	71.0	79.7	79.8
(5,6)	70.8	69.4	73.0	68.5	**74.0**
(2,8)	68.8	57.4	**76.5**	74.9	75.2
Fashion MNIST					
(0,1)	80.5	79.0	**84.9**	**84.9**	82.1
(2,8)	76.1	75.4	79.3	78.6	**79.5**
(3,9)	84.3	78.2	**87.9**	86.5	86.7
(7,9)	70.0	66.2	**68.4**	73.2	74.5
CIFAR					
(0,1)	60.4	59.0	**61.0**	57.8	58.5
PneumoniaMNIST					
(0,1)	70.5	63.4	**76.0**	67.3	68.7

Depolarizing noise was introduced at intensities of 3-10%

- Allows for comparison of baseline performance and resilience to noise
- Encoding techniques with lower EE **(ATP and SQE)** showed stronger robustness
    - Accuracy reductions were limited to **3-8 points**
- Angle, Amplitude, and PCA see accuracy drops between **4-17 points**

Adversarial Robustness

We applied Fast Gradient Sign Method (FGSM) with an attack strength of ε = 0.3 across datasets.



Attack effectiveness after adversarial training

- Direct encoding methods were more susceptible to attacks
- ATP, PCA, and SQE exhibit moderate robustness

After adversarial training was applied, **ATP and PCA had enhanced performance.**

- Filtering and pruning decreases sensitivity to irrelevant perturbations
- Further measures may be needed for robustness under higher ε

Conclusion

Limitations

- Selective pruning may introduce **bias toward high-variance regions**
  - Impact generalization
- Optimization incurs **moderate computational overhead**
  - Impact efficiency in larger applications

Future Work

- Explore **adaptability** to datasets with varied data distributions
  - Current evaluations focus on standard QNN benchmarks
  - Multi-class classification.

THANK YOU

Questions?

in/afane

mafane@fordham.edu