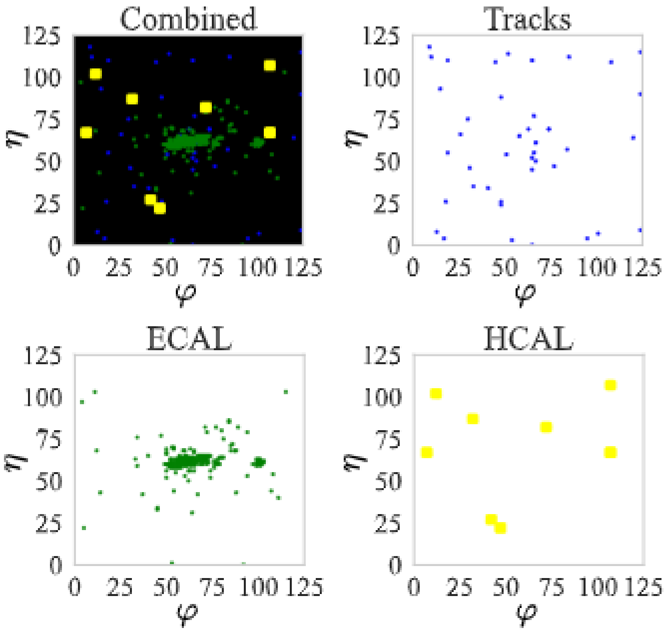
Quantum Attention for Vision Transformers in High Energy Physics

Alessandro Tesi   Sergei Gleyzer   Konstantin T. Matchev
Katia Matcheva   Kyoungchul Kong   Tom Magorsch
Gopal Ramesh Dahale

Introduction

- The **High Luminosity Large Hadron Collider (HL-LHC)** will generate vast amounts of data.
- **Quantum Machine Learning (QML)** and **Variational Quantum Algorithms (VQAs)** offer advantages in handling complex data.
- **Quantum Vision Transformers (QViTs)**: Integrate quantum circuits into **Vision Transformers (ViTs)** frameworks to improve efficiency and stability.



Methodology

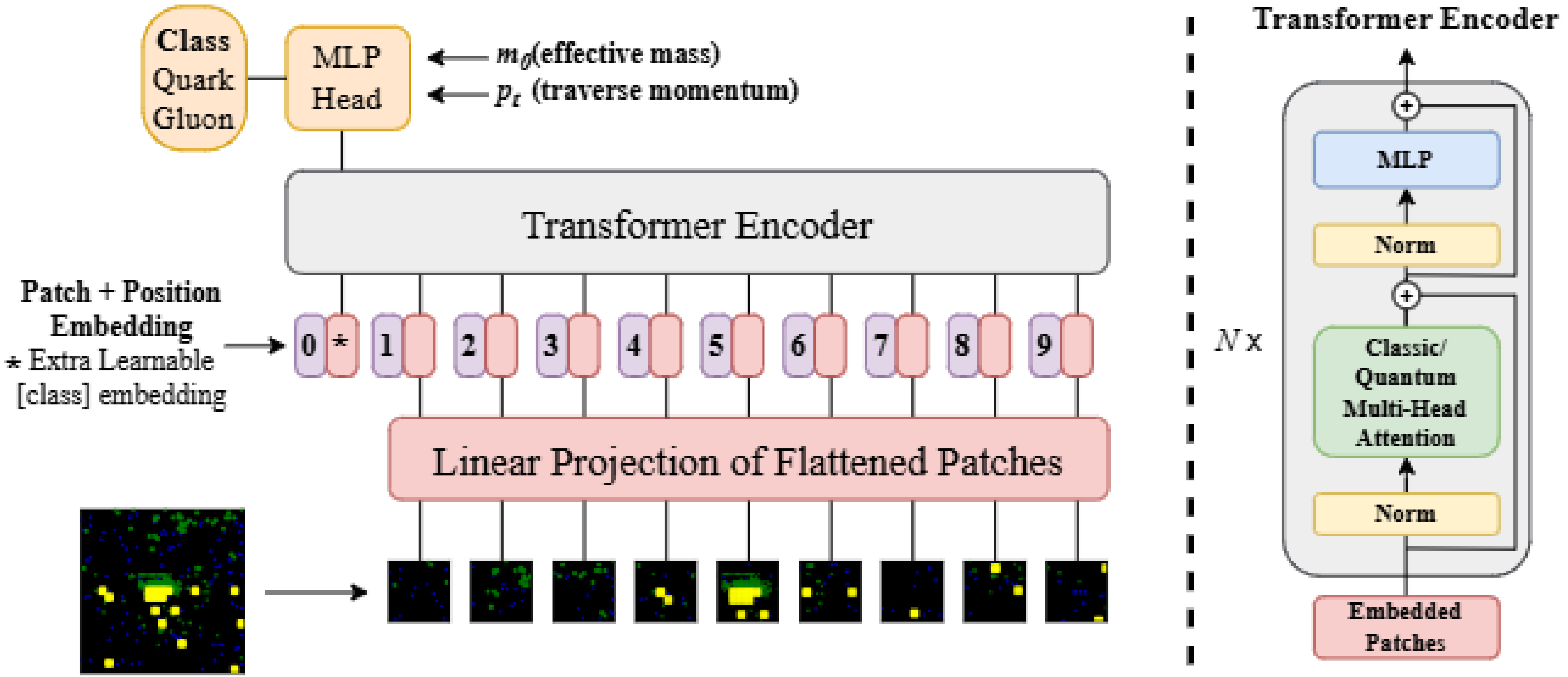
Transformer-based Architecture:

- Processes images as sequences of patches instead of convolutional layers.
- Uses Self-Attention Mechanism to capture long-range dependencies.

Pipeline:

- **Patch Extraction & Embedding**: Jet images are split into non-overlapping patches.
- **Multi-Head Self-Attention (MHSA)**: Computes attention scores between patches.
- **Feedforward Network (FFN)**: Processes refined patch embeddings for classification.
- **Classification Head**: Uses auxiliary jet features (transverse momentum pT and effective mass m0).

Methodology



Methodology

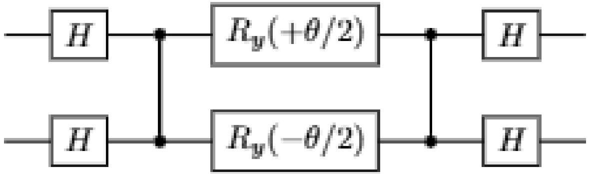
Reconfigurable Beam Splitter (RBS) Gates:

- Fundamental quantum gate used for orthogonal transformations.
- Implemented using Hadamard (H) gates, Controlled-Z (CZ) gates, and single-qubit Ry(±θ/2) rotations.

Vector Loading Circuits:

- Encode classical data into quantum states using unary amplitude encoding.
- Feature vector mapped to quantum state:

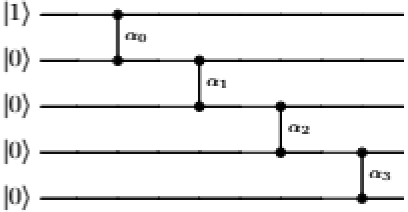
$$|\psi\rangle = x_0|10...0\rangle + x_1|01...0\rangle + ... + x_n|0...01\rangle$$



Decomposition of the RBS(θ) gate.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Rotation matrix applied in the two-dimensional subspace:



Vector loading circuit

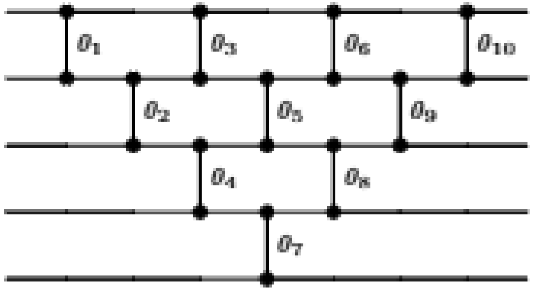
Methodology

Quantum Pyramid Circuits:

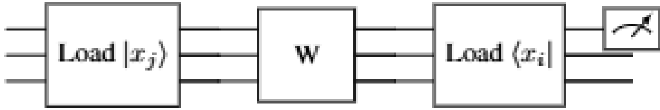
- Implements structured orthogonal transformations in quantum attention layers.
- Uses a pyramid of RBS gates to control quantum information flow.

Quantum Attention Coefficient Computation:

- Computes overlap between transformed query and key vectors using quantum measurement.
- Enables all attention computations by forming the attention map.



Example of a Pyramid Circuit



Circuit to compute an attention coefficient

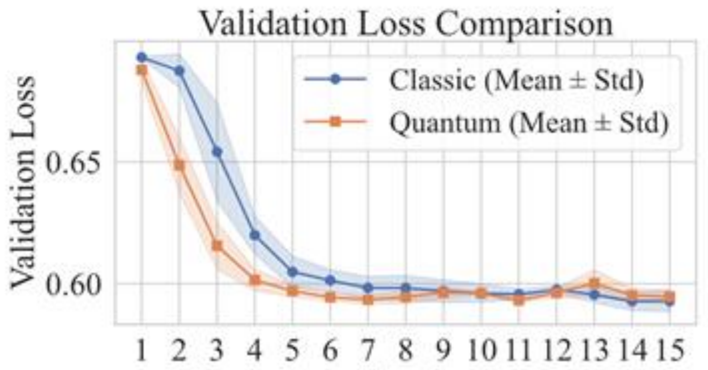
Results

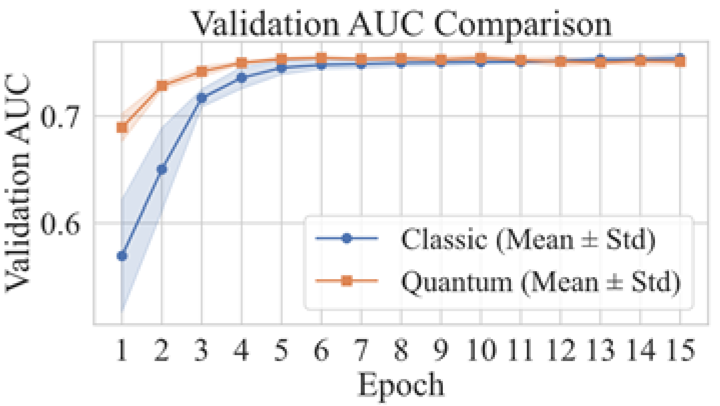
Dataset: CMS Open Data, with three subsets of 100,000 jet images each, 125x125 resolution, Tracks, ECAL, HCAL channels.

Training & Evaluation: 70% training, 15% validation, 15% test split.

Performance Metrics (Mean ± Standard Deviation):

- **Validation AUC:**
  - QViT: **0.749 ± 0.005**
  - Classical ViT: **0.751 ± 0.005**
- **Test AUC:**
  - QViT: **0.750 ± 0.006**
  - Classical ViT: **0.752 ± 0.006**





Conclusions

Key Takeaways:

- QViTs with QONNs maintain **robust classification performance**.
- **Quantum orthogonal transformations** enhance stability and computational efficiency.

Future Directions:

- **Hardware Implementation:** Test QViT on real quantum devices.
- **Quantum Particle Transformer for Jet Tagging:** Investigate a quantum-based approach for jet tagging with a Particle Transformer.