Extending a Classical Vision Transformer to a Quantum Vision Transformer

1. Introduction

A Quantum Vision Transformer extends the principles of a Classical Vision Transformer by using quantum computing for various stages of image processing and attention mechanisms. Quantum devices can provide exponential speedups for certain tasks, particularly in high dimensional data and optimization problems

---

2. Architecture

- Quantum Image Encoding : Classical image data is encoded into quantum states using quantum feature maps or amplitude encoding.

- Quantum Patch Embedding : Image patches are transformed into quantum embeddings using parameterized quantum circuits .

- Quantum Self-Attention : Quantum circuits implement a self-attention mechanism for contextual representation.

- Quantum Feedforward Network : A quantum enhanced feedforward network processes attention outputs.

- Hybrid Classifier : A combination of quantum circuits and classical neural networks performs the final classification.

---

3. Advantages of QViT

- Efficient High-Dimensional Processing : Quantum states can represent large-dimensional data efficiently.

- Enhanced Attention Mechanisms : Quantum entanglement enables complex relational reasoning.

- Hybrid Learning : Leveraging quantum computing and classical deep learning improves model flexibility.

---

4. Conclusion

Quantum Vision Transformers present a promising research direction. By integrating quantum circuits into the ViT architecture, QViTs can offer computational advantages in image processing tasks.