

Extending a Classical Vision Transformer to a Quantum Vision Transformer

1. Introduction

A **Quantum Vision Transformer (QViT)** is a more powerful version of a Vision Transformer (ViT). ViTs are great at analyzing images using patches, similar to how we focus on different parts of a picture. A QViT uses quantum computers to do this faster and more efficiently.

2. Architecture Overview

A QViT has these main parts:

- **Quantum Image Encoding:** It turns regular image data into quantum data that a quantum computer can understand.
 - **Quantum Patch Embedding:** Image patches are converted into quantum states using special circuits.
 - **Quantum Self-Attention:** This is where the quantum computer focuses on the most important parts of the image.
 - **Quantum Feedforward Network:** The results from the attention part are processed further using quantum circuits.
 - **Hybrid Classifier:** The final results are converted into a format a classical computer can read and classify.
-

3. Detailed Components

3.1. Quantum Image Encoding

- The image is split into small patches.
- Each patch is converted into quantum data using methods like amplitude encoding.

3.2. Quantum Patch Embedding

- Each patch goes through a quantum circuit that converts it into a quantum state.
- These circuits act like small functions that learn useful information from the image.

3.3. Quantum Self-Attention

- Quantum circuits are used to find which patches are the most important.
- They compare all patches and decide where to focus, just like how we look at key details in a picture.

3.4. Quantum Feedforward Network

- After finding the important parts, the quantum system processes this data further.

- It uses a network of quantum circuits to extract more insights.

3.5. Hybrid Classifier

- The quantum computer produces a result that is measured and converted into a classical format.
 - A traditional machine learning model can then use this result to make a final decision, like identifying an object in the image.
-

4. Advantages of QViT

- **Faster Processing:** Quantum computers can analyze large images faster.
 - **Better Attention:** Quantum systems can focus on the most important image features efficiently.
 - **Efficient Data Use:** They handle large image datasets using fewer resources.
-

5. Conclusion

Quantum Vision Transformers are a new and exciting way to analyze images using the power of quantum computing. While they are still being developed, they have the potential to improve image recognition in fields like medicine, security, and autonomous driving.