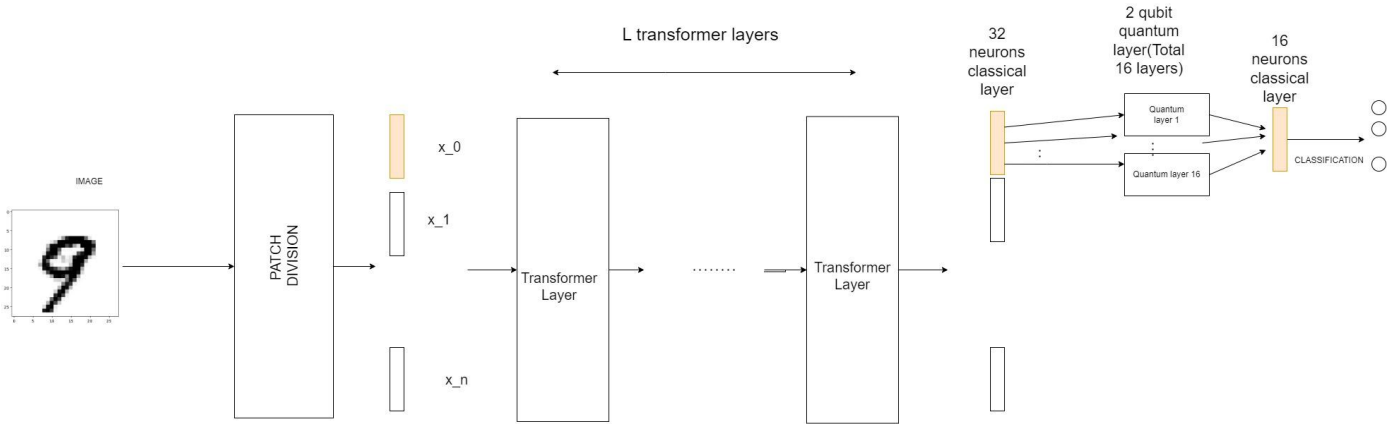
Extending the classical vision transformer to Quantum Vision Transformer.

I propose 3 different architectures for extending vision transformer to quantum vision transformer

Architecture-1) FeedForward Quantum Architecture-

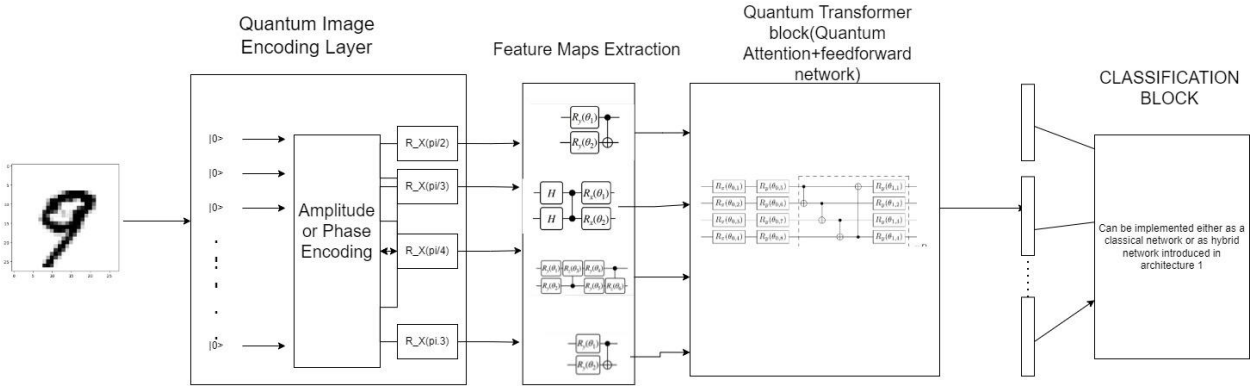


The above architecture is quite straightforward and very simple. Until the *final feed forward network* we implement the architecture exactly the same as our normal vision transformer.

In the final feed forward network ,if we consider it to have 32 neurons then we connect the first two neurons of the classical layer to quantum layer 1 consisting of 2 qubits, next two neurons of classical layer to quantum layer 2 and so on, the last 2 neurons of classical layer are connected to the last 2 qubits in the quantum layer 16.

NOTE:- If the computational cost is too expensive we can reduce the number of final quantum layers to 8 or 4 by projecting the feedforward network first to a dimension of 16/8. In this manner we can manage the computational cost in the network .

Architecture-2) Hybrid Vision Transformer-



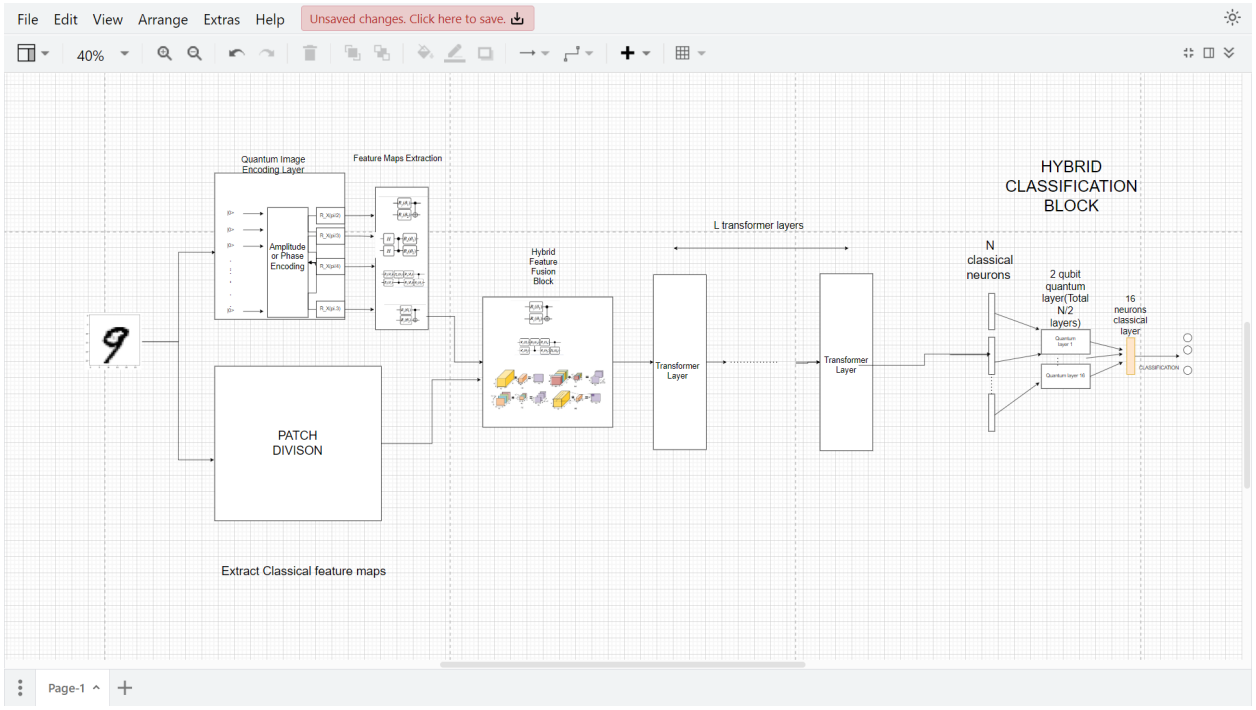
Here in this architecture at the beginning we will use Quantum Image Encoding or Quantum Patch Embedding instead of our normal classical patch embedding and it will consist of various quantum encoding techniques like Amplitude Encoding,Phase Encoding or quantum fourier transforms etc.

Then we pass it through a feature map extractor to extract the features using a set of parametrized quantum circuits like the convolution layers presented in the Quantum convolutional neural network for classical data classification paper [1].

After this we pass our feature maps through the Quantum TransformerBlock which will consist of the QuantumAttention and QuantumFeedForward Network/Classical Feedforward network.

Once this process is done we pass the quantum output to the Quantum Layer as illustrated in the Architecture-1 and then pass it to the classical feedforward network and then do the classification.

ARCHITECTURE 3- Hybrid Multi-Feature Transformer-



The above architecture uses both Quantum and Classical feature map extractor implemented in architecture 1 and 2 to extract features.

Once both the feature maps are extracted they are passed through Hybrid_Feature_Fusion Block that takes both quantum and classical feature maps as input which enables it to be able to capture both quantum and classical information about the image.

Now this hybrid captured information is passed through the transformer block enabling it to capture long range dependencies between image features.

Now finally this output is classified using feedforward_quantum_architecture(aka Architecture 1).

References-

1).Quantum convolutional neural network for classical data classification
2).Quantum Vision Transformers
3).THE DAWN OF QUANTUM NATURAL LANGUAGE PROCESSING