HALO: High Autonomous Low-SWaP Operations

User/Developer Manual

Sloan Hatter, Blake Gisclair
Dr. Ryan T. White

Florida Institute of Technology

April 20, 2026

System Architecture Diagram

Figure 1 depicts the system architecture diagram for a Vision Transformer (ViT) neural network. A ViT is a type of neural network architecture that applies the transformer model to image data. The transformer breaks an image into patches and processes the sequence of patches at once using attention to connect related image pieces, allowing it to learn which parts of the input are important to each other. One important aspect of ViTs is that during the image patching step, positional embeddings are added to each patch so order and spatial information of the image is preserved; this means that ViTs essentially know where each patch goes within an image, allowing it to retain global context.

Parts of a Vision Transformer:

1. Image to Patch Extraction
2. When an image is fed into a transformer, it is split up into fixed-size patches, producing a grid of patches, as transformers need sequences

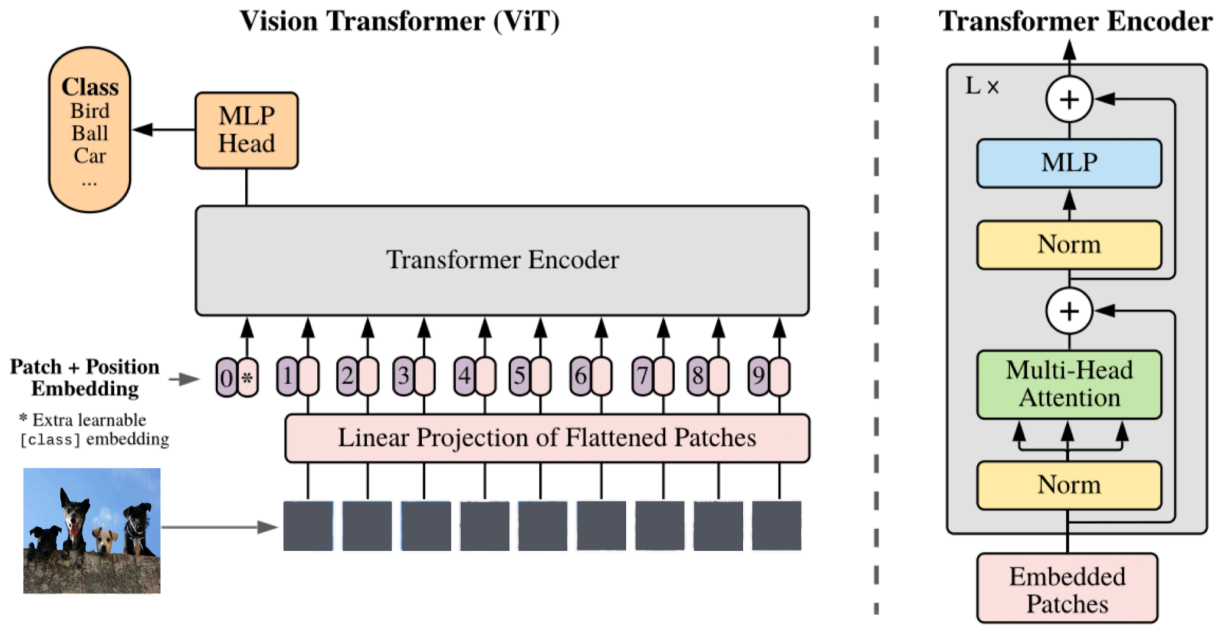

Figure 1

Source Files

Quantization