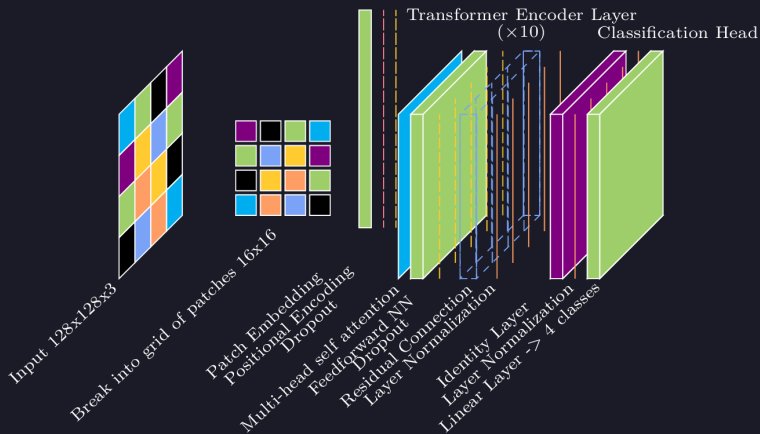# VIT Architecture

Number of parameters: 21459460

Training Details

VIT training and model parameters:
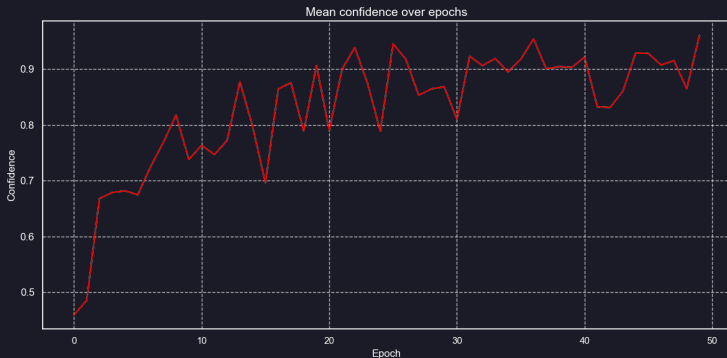
- **Epochs**: 50
- **Optimizer**: Adam (weight decay 1×10^{-5})
- **Scheduler**: stepLR (step size 10, gamma 0.5)
- **Loss function**: Cross-entropy
- **Learning rate**: 1×10^{-4}
- **Batch size**: 64 (both training and validation)
- **Activation function**: Mish
- **Dropout rate**: 0.2
- **Image size** and **Patch size**: 128×128, 16×16
- **Number of heads**: 8
- **Number of layers**: 10
- **Patch embedding dimension**: 512
- **Feedforward dimension**: 1024

Training Loss and Accuracy

- Final training loss: 0.27
- Final training accuracy: 90%

Confidence and Test Accuracy

- Final training confidence: 96%
- Final test confidence: 93%
- Final test accuracy: 88%