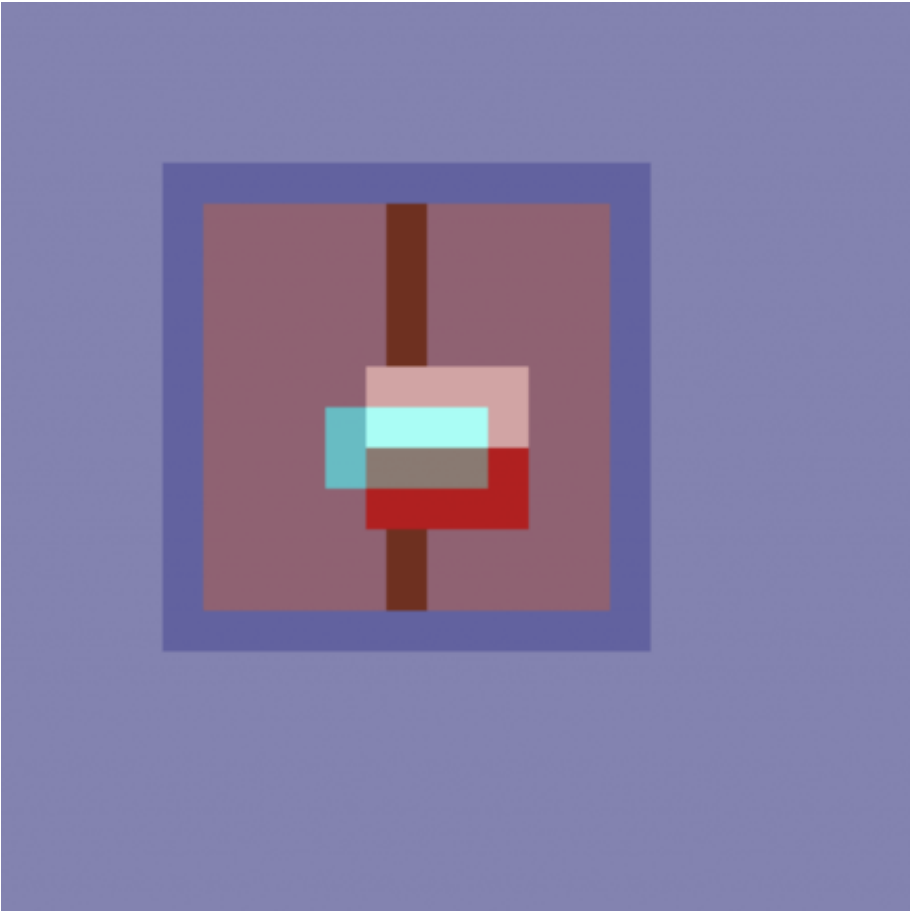


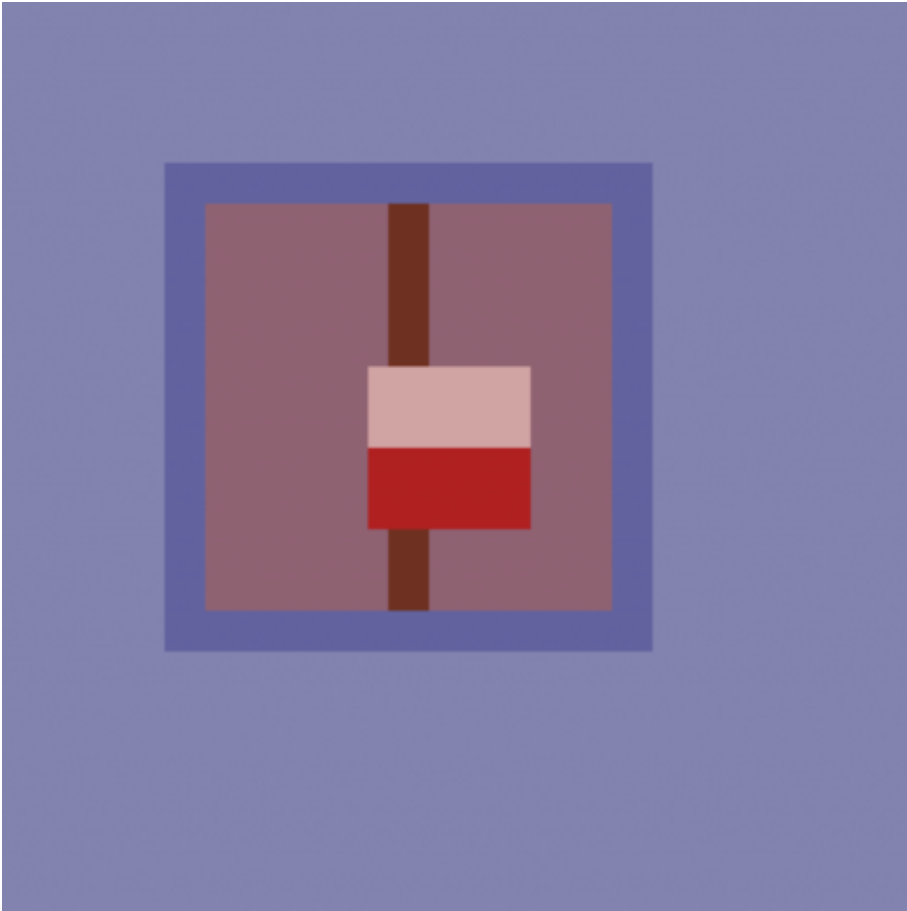
Attention Pattern Analysis in Flag Classification

Standard CLIP
Accuracy: 18.0%



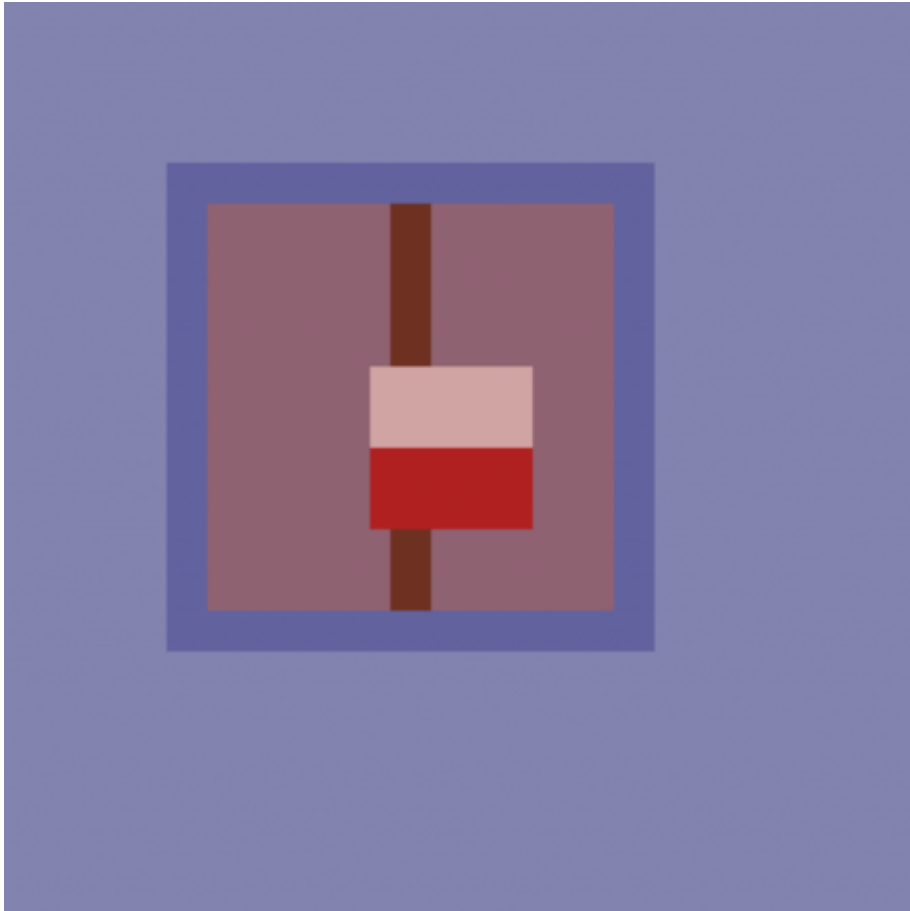
- Focuses on building (60%)
- Minimal flag attention (15%)
- Poor symbol recognition

RS5M Fine-tuned (artifact: majority-class collapse)
Accuracy: 72.6%



- Strong flag focus (70%)
- Symbol recognition improved
- Limited context understanding

RS5M + Hierarchical
Accuracy: 94.8%



- Hierarchical attention
- Symbol + context balance
- Optimal performance

