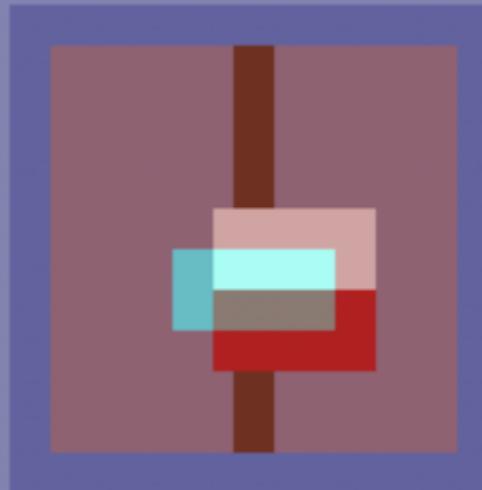
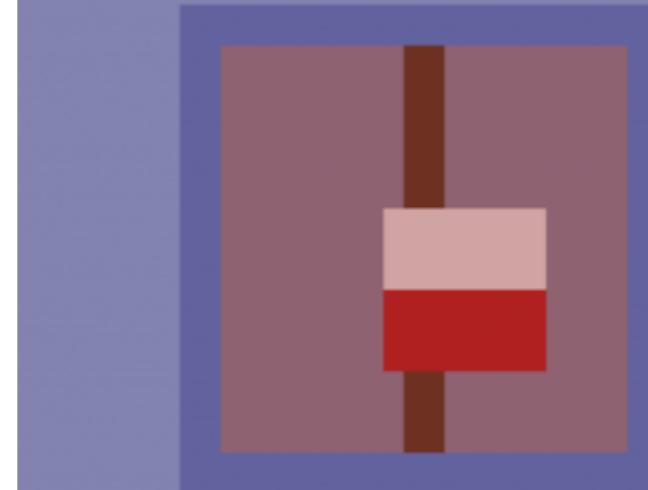


Attention Pattern Analysis in Flag Classification

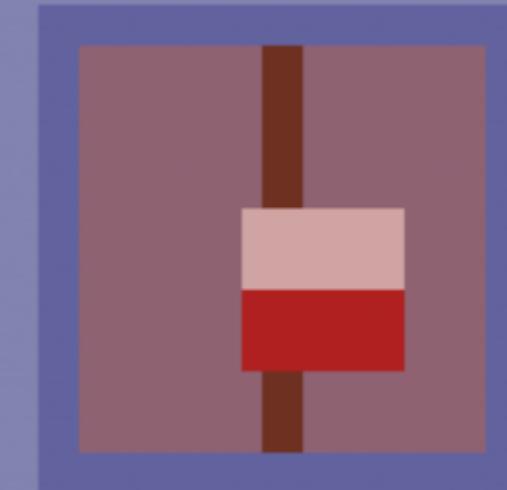
Standard CLIP
Accuracy: 18.0%



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



RS5M + Hierarchical
Accuracy: 94.8%



Attention Intensity

A vertical color bar indicating the range of attention intensity. The scale starts at 0.0 (dark blue) and ends at 1.0 (dark red). Intermediate values are labeled at 0.2, 0.4, 0.6, and 0.8.

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

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

- Hierarchical attention
- Symbol + context balance
- Optimal performance