Pre-Training

Hierarchical Transformer

~50x50

~20

M0
M1
M2
M3
M4
M5
M6
M7

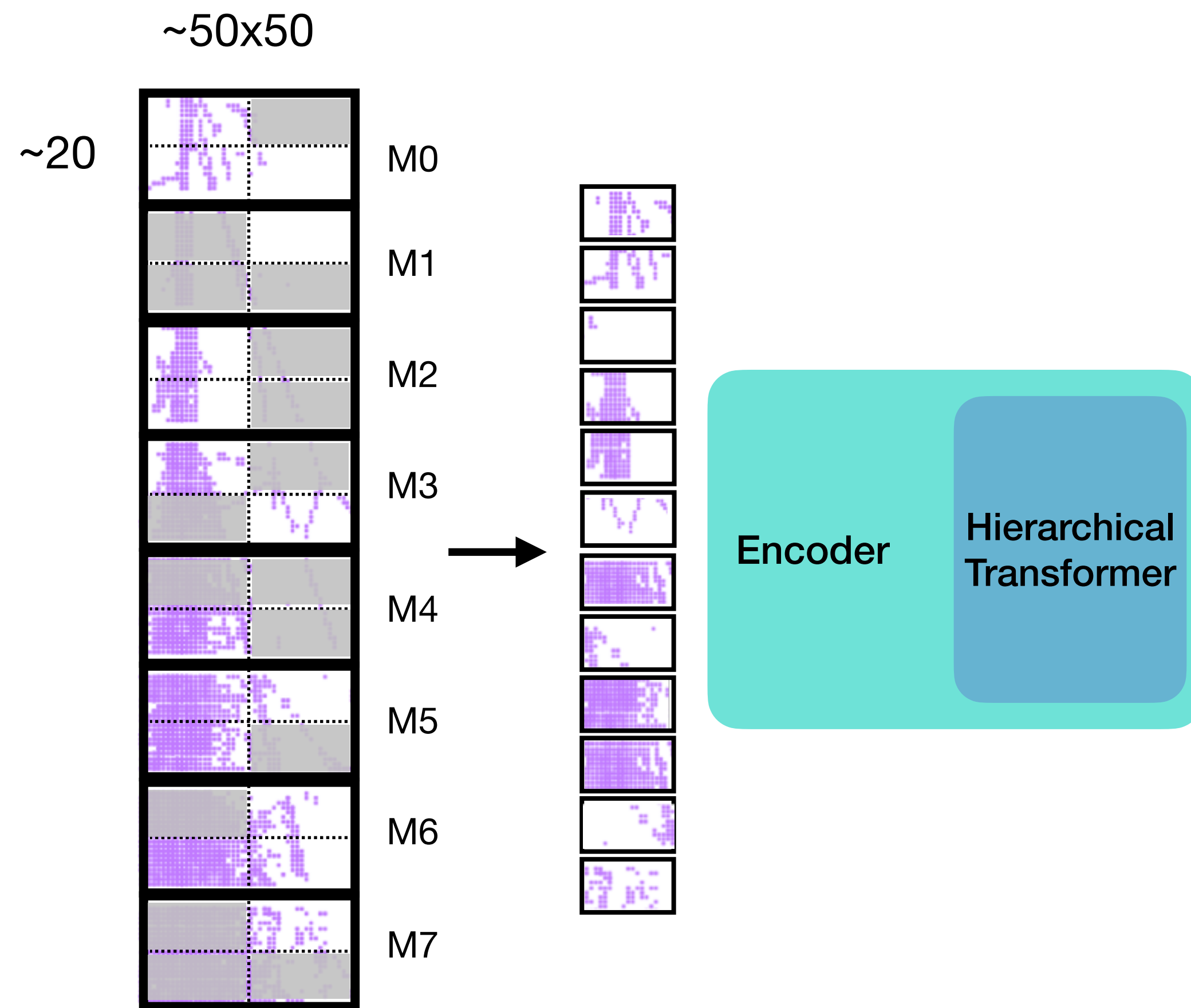Encoder | Hierarchical Transformer

- **Intra-Module Attention (The Local View):** Summarize hits *within* each detector module independently

- **Inter-Module Attention (The Global View):** Combine module summaries → learn long-range relations & event topology. Also uses global information form other sub detectors (RealCal, HCal, MuTag)

768 features → 768 features

Embedding of one patch enriched of Local and Global info

51