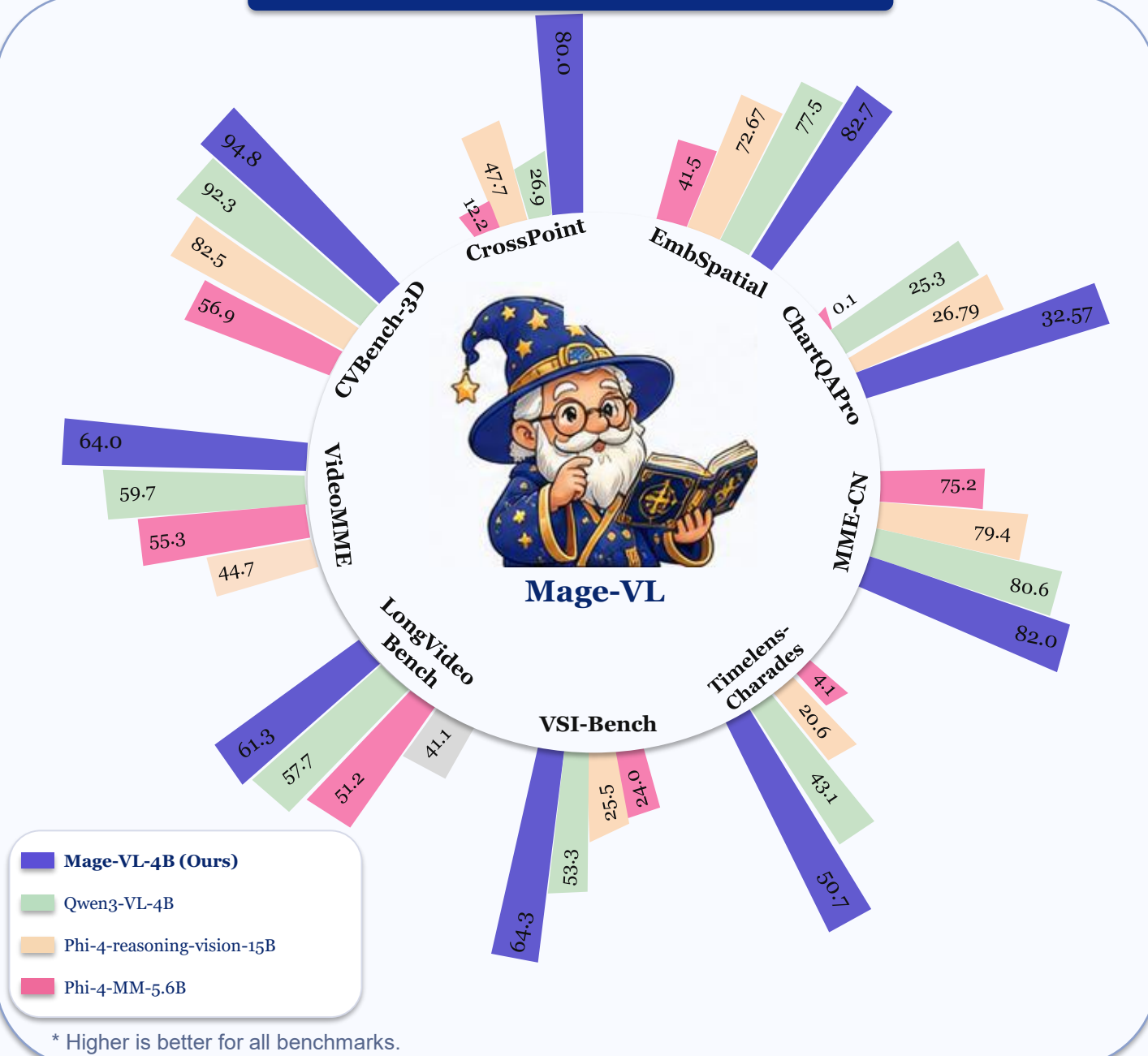




Mage-VL

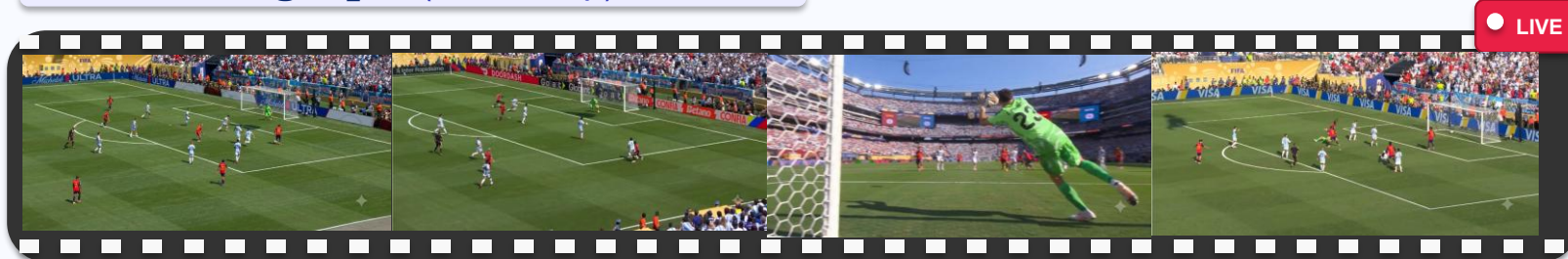
Codec-Native Visual Representations
for Efficient Multimodal Understanding

Benchmark Performance

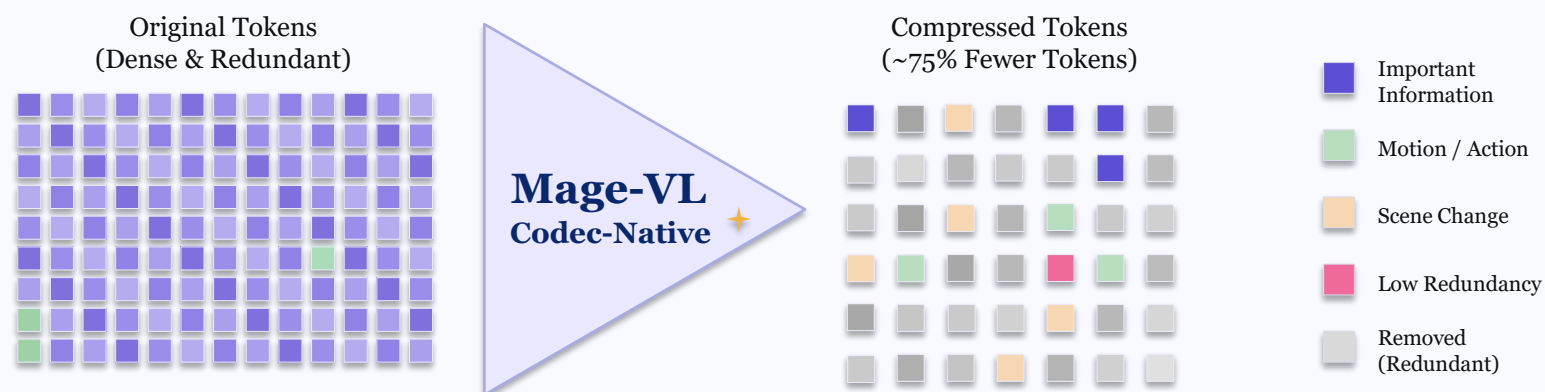


(Streaming) Video Understanding

1. Streaming Input (World Cup)



2. Codec-Native Token Selection



3. Online Multimodal Understanding

