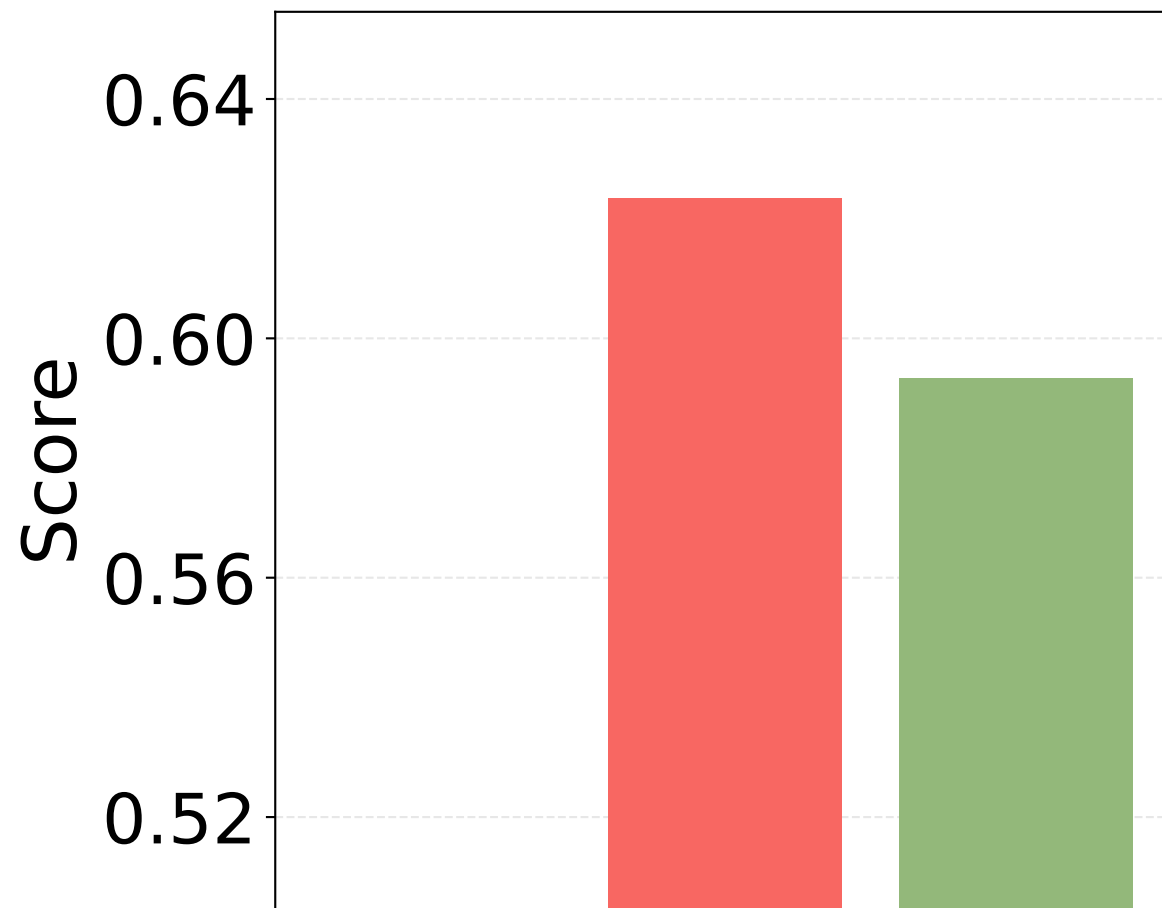
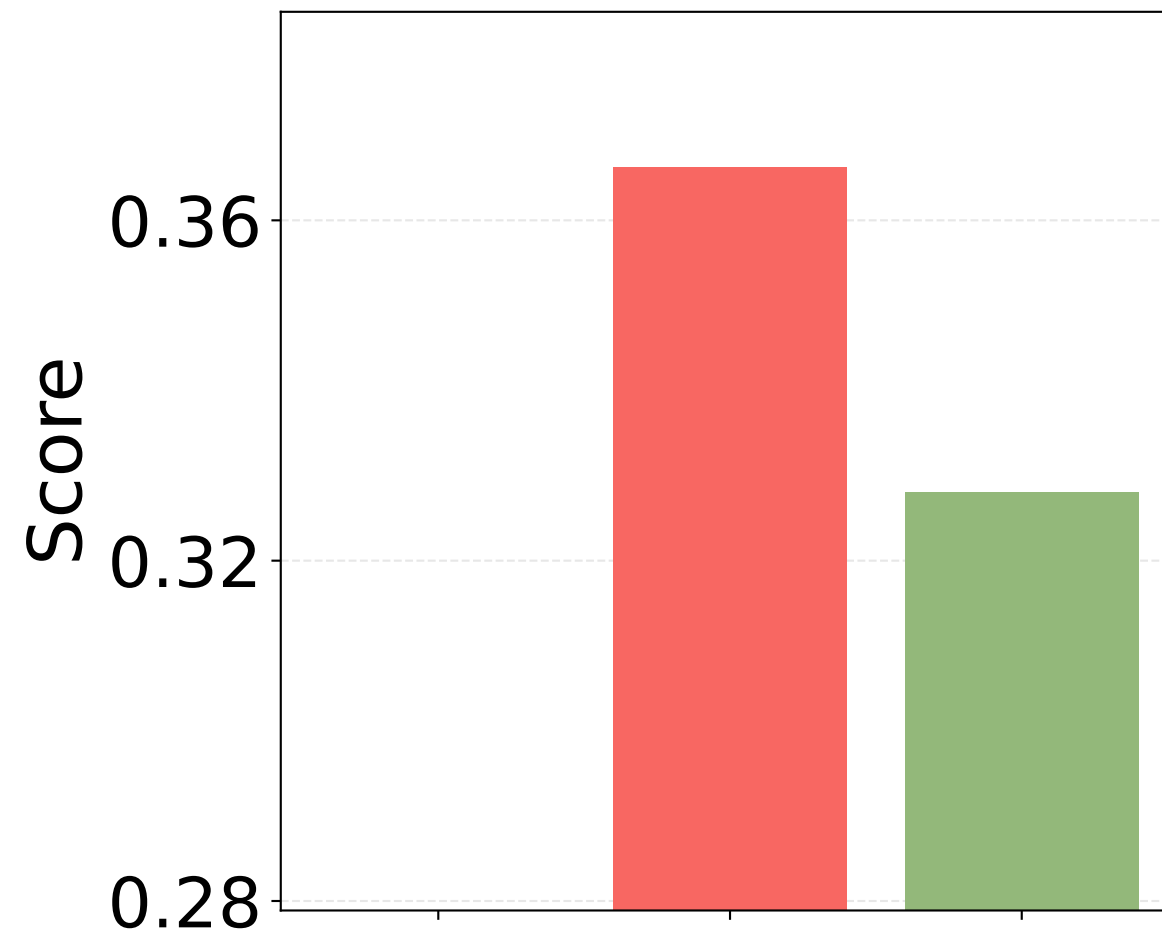
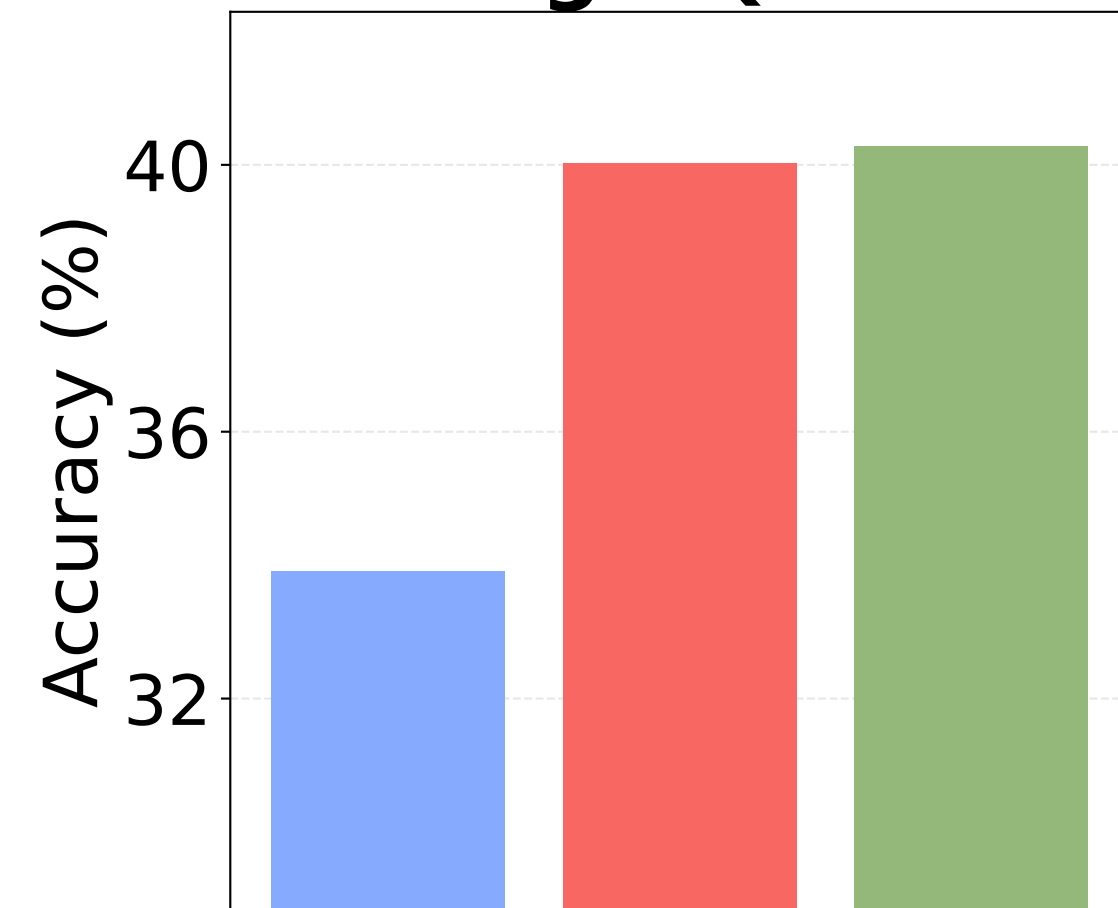


DPGBench ↑**GenEval ↑****Avg VQA ↑**

520B Text + 520B Multimodal Tokens



Text + Video



Text + MetaCLIP



Text + Video + MetaCLIP + Action