Static Benchmarks Static Benchmarks Evolving Better LMM Result Evaluation Evaluation Result

Fully Dynamic Multimodal Evaluation Framework

Enabling benchmarks to keep pace with rapidly evolving LMM



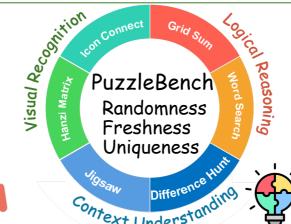
Diverse Elements Open-ended Rules Unique VQA

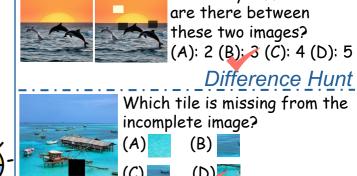
Open-ended Visual Puzzle Generation(OVPG)

Overview of PuzzleBench

Focusing on 3 Dimensions with 6 Puzzles, Total 12K VQA Samples







Jigsaw