



(a) Previous Benchmarks: Each question paired with a limited image set

(b) Our Benchmarks:
All questions mapped to an extensive document collection