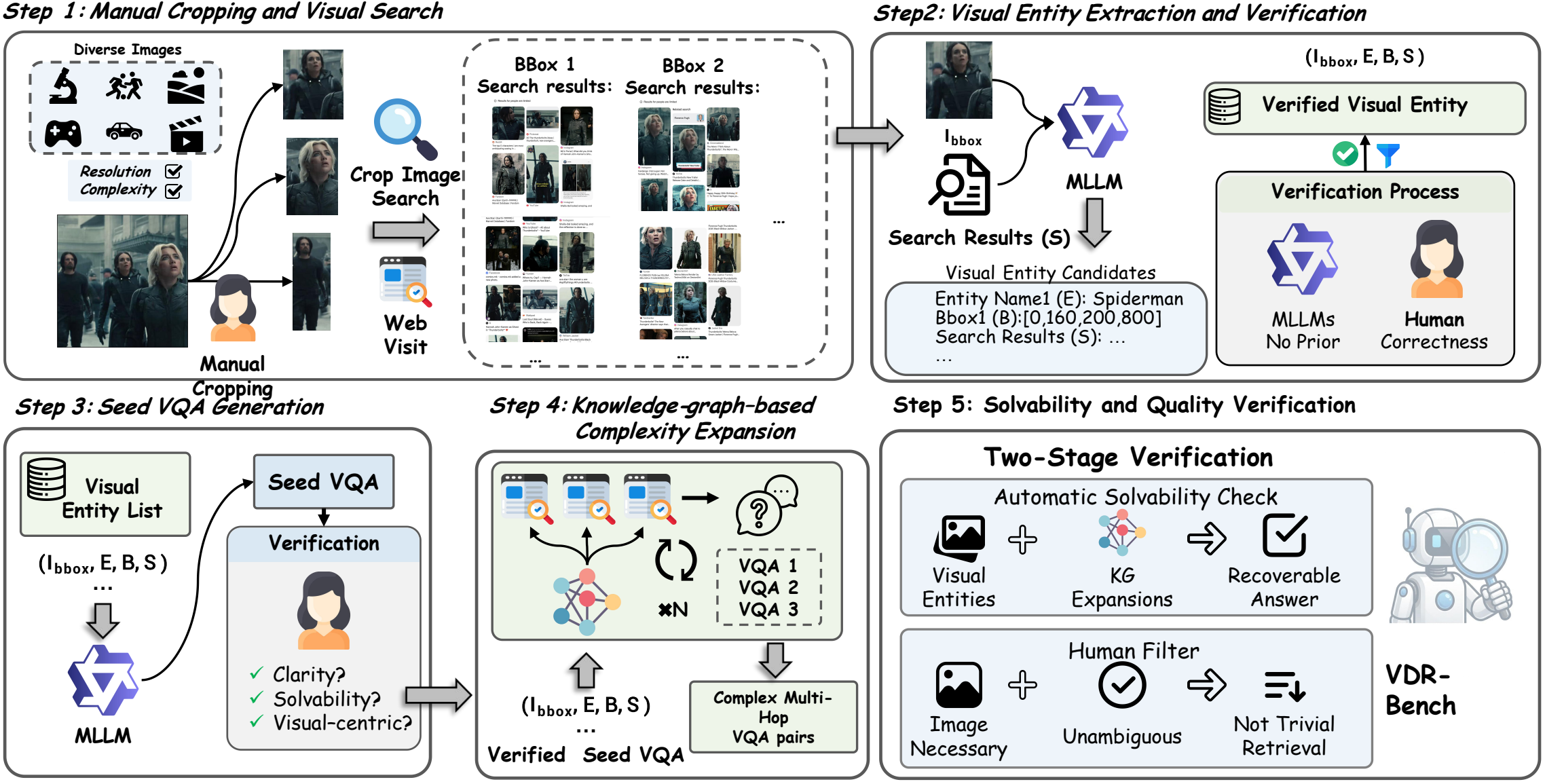
Step 1: Manual Cropping and Visual Search
Diverse Images
Resolution
Complexity
Manual Cropping
Crop Image Search
Web Visit
BBox 1 Search results:
BBox 2 Search results:
Step2: Visual Entity Extraction and Verification
I_{bbox}
MLLM
Search Results (S)
Visual Entity Candidates
Entity Name1 (E): Spiderman
Bbox1 (B):[0,160,200,800]
Search Results (S): ...
...
(I_{bbox}, E, B, S)
Verified Visual Entity
Verification Process
MLLMs No Prior
Human Correctness
Step 3: Seed VQA Generation
Visual Entity List
(I_{bbox}, E, B, S)
...
MLLM
Seed VQA
Verification
Clarity?
Solvability?
Visual-centric?
Step 4: Knowledge-graph-based Complexity Expansion
VQA 1
VQA 2
VQA 3
×N
(I_{bbox}, E, B, S)
...
Verified Seed VQA
Complex Multi-Hop VQA pairs
Step 5: Solvability and Quality Verification
Two-Stage Verification
Automatic Solvability Check
Visual Entities
KG Expansions
Recoverable Answer
Human Filter
Image Necessary
Unambiguous
Not Trivial Retrieval
VDR-Bench