1. Scene Splitting

Video Datasets

2. Extract Frames

Selected ✅ High similarity to the left frame ❌ Low Sharpness ❌

Filter Criteria
- Sharpness
- Similariy

3. Object Extraction

Person Caption

Animal Caption

...

Object Detection & Caption

7. Human Verification

Annotation Criteria

- MTQA Accuracy Checking
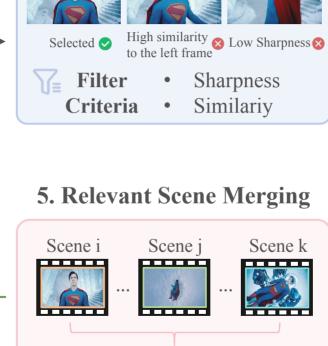- Task Quality Verification
- Contextual Coherence
- ......

6. MTQA Generation

Scene Settings
- Single Scene
- Cross Scene

6 Core Abilities
- Perceptivity
- Interactivity

5. Relevant Scene Merging

Scene i   Scene j   Scene k

Object-Based Scene Merging

4. Object Memory Bank

Time

ID: ID 1
Information: ......

ID: ID 2
Information: ......

......