

1 Scene Detection (split input into scenes)

Reordering

3 Visual Processing

4 Dialogue Summarization

5 High-level Summarization