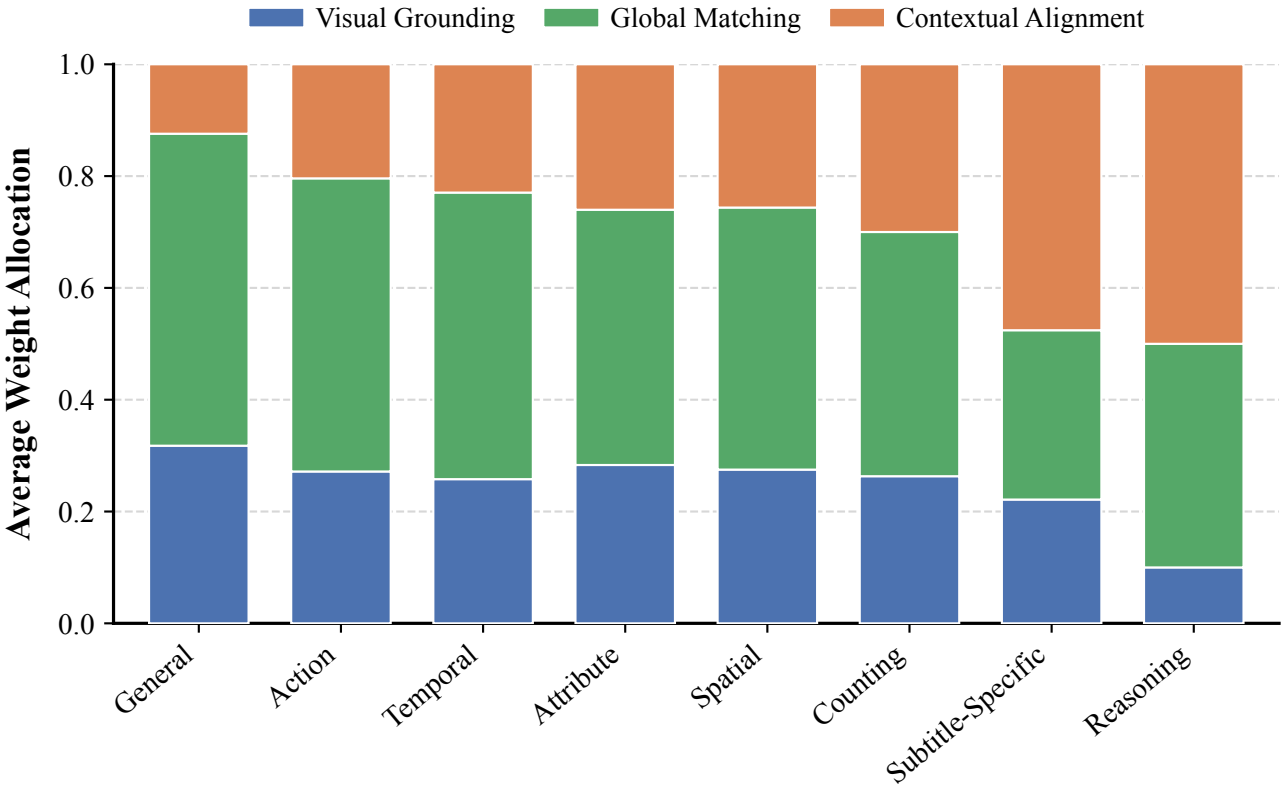

Visual Grounding
Global Matching
Contextual Alignment
Average Weight Allocation
1.0
0.8
0.6
0.4
0.2
0.0
General
Action
Temporal
Attribute
Spatial
Counting
Subtitle-Specific
Reasoning