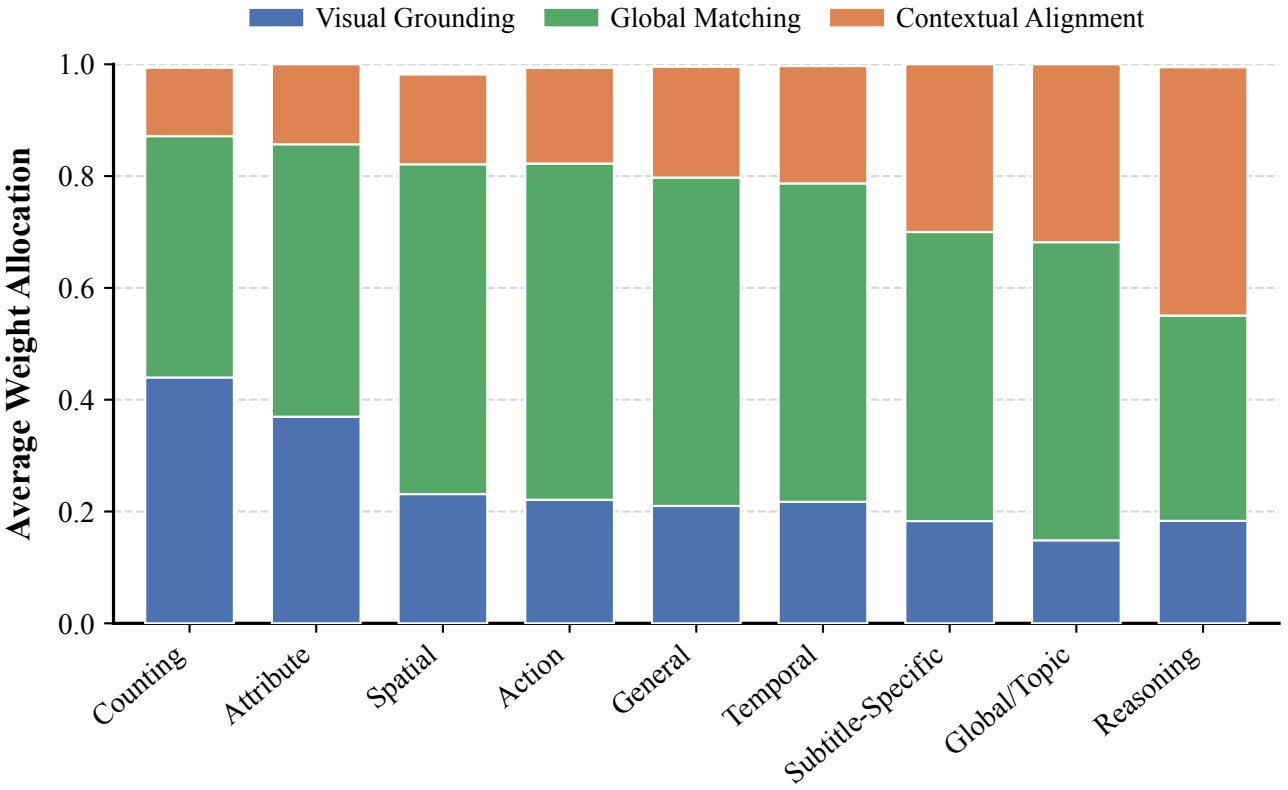

Visual Grounding
Global Matching
Contextual Alignment
Average Weight Allocation
1.0
0.8
0.6
0.4
0.2
0.0
Counting
Attribute
Spatial
Action
General
Temporal
Subtitle-Specific
Global/Topic
Reasoning