

Return

4%

Intra-modal attentions

Cross-modal attentions

Action

17%

10%

State

20%

23%

26%

Return

Action

State

