1. Beam Search Algorithm

2. Dataset

Fig. 1 showcase some snapshots from our dataset. Every three rows correspond to one long video, wherein the first row is the third-person view, and the other two rows are the first-person views from two agents. The first video is mainly about Joint Attention. The second video includes No Communication, Attention Following and Joint Attention; it also involves second-order false belief. The third video includes Attention Following. The fourth video includes No Communication.

3. Surveys for Human Studies

Below are the links to the questionnaires for the human subject studies in the keyframe-based video summary task.

• Group 1: https://5minds.typeform.com/to/dh782Z
• Group 2: https://5minds.typeform.com/to/T3hGhN
• Group 3: https://5minds.typeform.com/to/wovakS
• Group 4: https://5mind.typeform.com/to/SpOMu3

4. Additional Quantitative Results

4.1. ROC curve

Fig. 3 show the ROC curves for all five minds in the predicting belief dynamics task. The numbers of belief dynamics denote different categories: 0–occur, 1–disappear, 2–update, and 3–null.

5. Additional Qualitative Results

Fig. 2 shows additional qualitative results for the keyframe-based video summary task.

*Lifeng Fan and Shuwen Qiu contributed equally.
Figure 1: Sample snapshots of the Meditation dataset.
Figure 2: Additional comparisons on video summarization.
Figure 3: ROC Curve