We sincerely appreciate the reviewer's recognition of our work and the increased score.

Regarding the Latent Space Temporal Interpolation module, we have decided to remove the 'logical' word from the paper to avoid confusion. It also should be noted that the module is not designed for concrete but implicit logical and casual inference.

For the comparison with state-of-the-art methods, we would like to address several points:

- First, regarding the comparison with SCHEMA[1], their use of Large Language Models (LLMs) introduces external knowledge not present in the original dataset, making it an unfair comparison.
- Second, for COIN with T=3, our analysis in Table 9 (line 843) demonstrates that increasing model size alone can improve performance. This finding suggests that for larger datasets, it is essential to develop a memory mechanism to store relevant information, rather than focusing solely on temporal relationships. We will explore this direction in our future research.
- Finally, our analysis in Table 15 (line 1012) shows that for T=6 compared to T=5, there is a larger gap between real and interpolated features. This suggests our current interpolation strategy may be insufficient for longer sequences, as features lack diversity to capture actions fully. We will explore improving interpolated feature quality in future work.

[1] Niu, Y., Guo, W., Chen, L., Lin, X., & Chang, S. F. (2024). SCHEMA: State CHangEs MAtter for Procedure Planning in Instructional Videos. arXiv preprint arXiv:2403.01599.