Generative Video Diffusion for Unseen Novel Semantic Video Moment Retrieval

Dezhao Luo[1], Shaogang Gong[1], Jiabo Huang[2], Hailin Jin[3], Yang Liu[4,5]
[1]Queen Mary University of London, [2]Sony AI, [3]Adobe Research,
[4]WICT, Peking University, [5]State Key Laboratory of General Artificial Intelligence, Peking University,

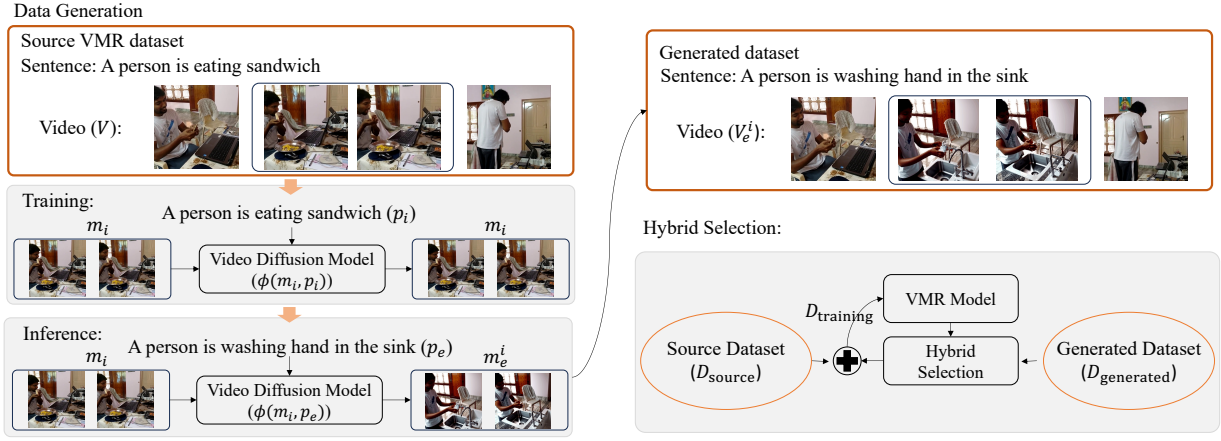
Problem Definition

The correlation between video moments and text is crucial for **video moment retrieval**, yet there is a scarcity of large-scale datasets.

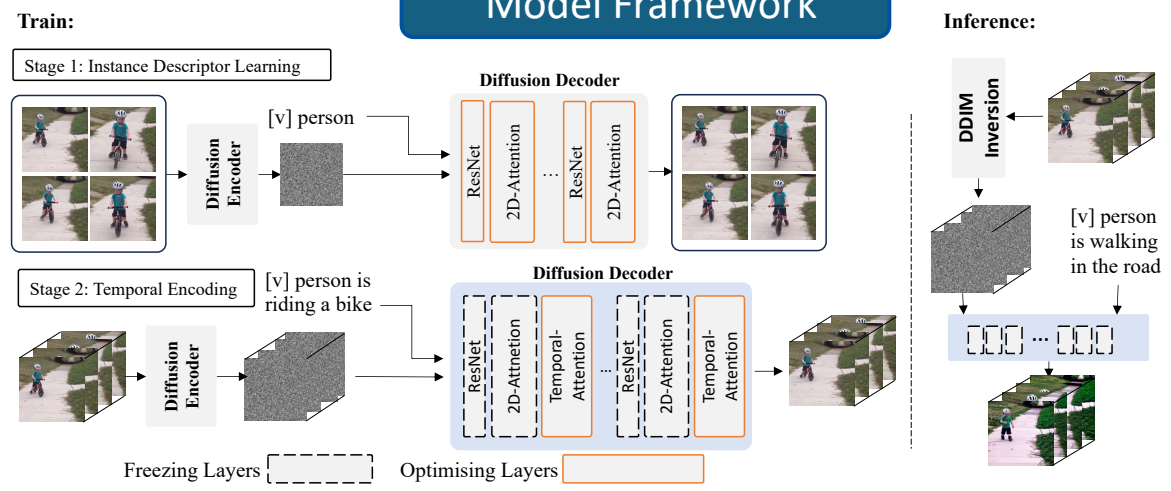
Solution

- A video diffusion model that synthesises training data
- A data selection module that selects beneficial data for the VMR task

Data Generation Pipeline



Model Framework



Video Editing Ability

Conclusion

- FVE is able to generate training data that benefits the VMR task
- FVE is able to change the action in a video and maintain other details.