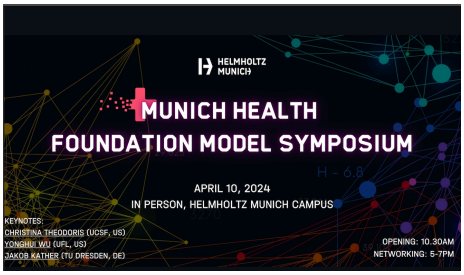

Contribution ID: **36** Type: **Poster**

EgoCVR: An Egocentric Benchmark for Fine-Grained Composed Video Retrieval

In Composed Video Retrieval, a video and a textual description which modifies the video content are provided as inputs to the model. The aim is to retrieve the relevant video with the modified content from a database of videos. In this challenging task, the first step is to acquire large-scale training datasets and collect high-quality benchmarks for evaluation. In this work, we introduce EgoCVR, a new evaluation benchmark for fine-grained Composed Video Retrieval using large-scale egocentric video datasets. EgoCVR consists of 2,295 queries that specifically focus on high-quality temporal video understanding. We find that existing Composed Video Retrieval frameworks may not achieve the necessary high-quality temporal video understanding for this task. To address this shortcoming, we adapt a simple training-free method, propose a generic re-ranking framework for Composed Video Retrieval, and demonstrate that this achieves strong results on EgoCVR.

Primary author: Mr HUMMEL, Thomas

Co-authors: Dr GEORGESCU, Mariana-Iuliana; KARTHIK, Shyamgopal (University of Tübingen); AKATA, Zeynep (Helmholtz Munich)

Session Classification: Break + Posters session

Track Classification: Poster