Fusion Methods are needed to be specified because we have 3 form of data with different modalities

Feature-level fusion

Decision-level fusion

Hybrid Fusion

Transformer – Cross-attention

Dataset

CMU-MOSI Dataset

The Multimodal Corpus of Sentiment Intensity (CMU-MOSI) dataset is a collection of 2199 opinion video clips. Each opinion video is annotated with sentiment in the range [-3,3]. The dataset is rigorously annotated with labels for subjectivity, sentiment intensity, per-frame and per-opinion annotated visual features, and per-milliseconds annotated audio features.

<http://multicomp.cs.cmu.edu/resources/cmu-mosi-dataset/>

CMU-MOSEI Dataset

CMU Multimodal Opinion Sentiment and Emotion Intensity (CMU-MOSEI) dataset is the largest dataset of multimodal sentiment analysis and emotion recognition to date. The dataset contains more than 23,500 sentence utterance videos from more than 1000 online YouTube speakers. The dataset is gender balanced. All the sentences utterance are randomly chosen from various topics and monologue videos. The videos are transcribed and properly punctuated

<http://multicomp.cs.cmu.edu/resources/cmu-mosei-dataset/>

<https://github.com/A2Zadeh/CMU-MultimodalSDK>

Other Dataset

