

Supplementary Materials: OpenAVE: Moving towards Open Set Audio-Visual Event Localization

Anonymous Authors

Table 1: Audio-Visual Event Dataset Splits for Open Set AVE. For each split, Seven out of twenty-eight event classes are randomly selected as the unknown (U) utilized in open set testing, while the rest categories are the known (K) in model training.

Class	Split1	Split2	Split3
Man speaking	K	K	K
Mandolin	K	U	K
Church bell	K	K	U
Rodents	K	K	K
Bark	U	U	K
Airplane	K	K	U
Race car	U	K	K
Female speech	K	K	U
Helicopter	U	K	K
Violin	K	K	K
Flute	K	U	K
Ukulele	K	K	K
Frying	U	K	K
Truck	K	U	K
Shofar	K	K	K
Motorcycle	K	K	K
Acoustic guitar	U	K	K
Train horn	K	K	K
Clock	K	K	U
Banjo	K	K	U
Goat	K	U	K
Baby cry	U	K	U
Bus	K	K	K
Chainsaw	K	K	K
Cat	U	K	U
Horse	K	U	K
Toilet flush	K	K	K
Accordion	K	U	K

1 OPENAVE DATASET DETAILS

To adapt the existing audio-visual event localization dataset (AVE dataset) for the open set audio-visual event localization setting, a subset of event classes need to be reserved as the unknown utilized in the open testing stage. To be specific, we randomly divide AVE dataset three times into known and unknown subsets of event classes. For each dataset split, the model is trained on its known events set, and evaluated on the open set containing known and unknown event categories. Table 1 shows detailed information of three splits from AVE dataset.