Dataset	Data type	Annotation type	Usage in this paper
YouTube-8M	YouTube video	Video-level label	Training of the sound
YT8M-IPA	YouTube video	Playing-action key point we annotate	Evaluation of the acti
AudioSet	YouTube video	Video-level sound label	Training of the sound
MedleyDB	Music audio	Frame-level sound label	Evaluation of the sou
ImageNet-Instrument[1]	Image	Instrument bounding box	Evaluation of the obj
MagnaTagATune	Music audio	Track-level music-related label	Training of the sound