

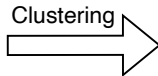
Canine Vocalization Extraction (Denoise)



Canine Vocalization Segmentation



Audio Frame Encoder



Canine Vocalization Transcription
42 [32, 5] [20, 32, 20] 31 [49, 21] [48, 5]



Parsing

Row Transcription
42 32 5 20 32 20 31 49 21 48 5



Replacement

