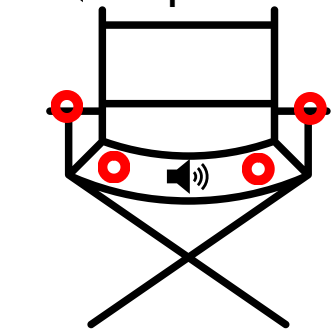
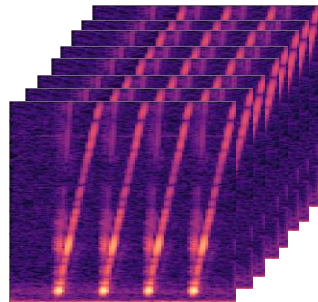


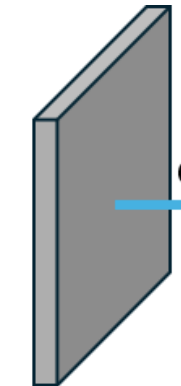
○ : Microphones,  
🔊 : Speaker



Active Acoustic  
Sensing



Spectrogram



Conv



$254 \times 58 \times 12$

Conv



$125 \times 27 \times 16$



54000

FC



1000

FC



12



Prediction

Pose Classification Network