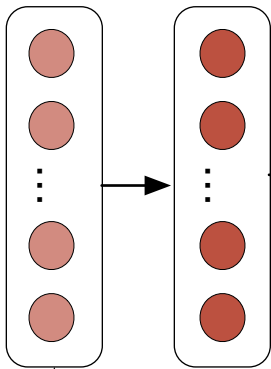
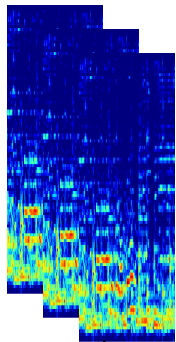
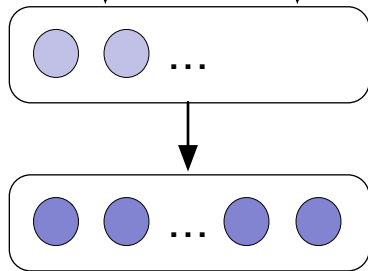


Audio input



Shared Layers



Prediction

Video input

