#1 Pre-training

Pre-training data:

- annotated (supervised)
- not annotated (unsupervised / self-supervised)

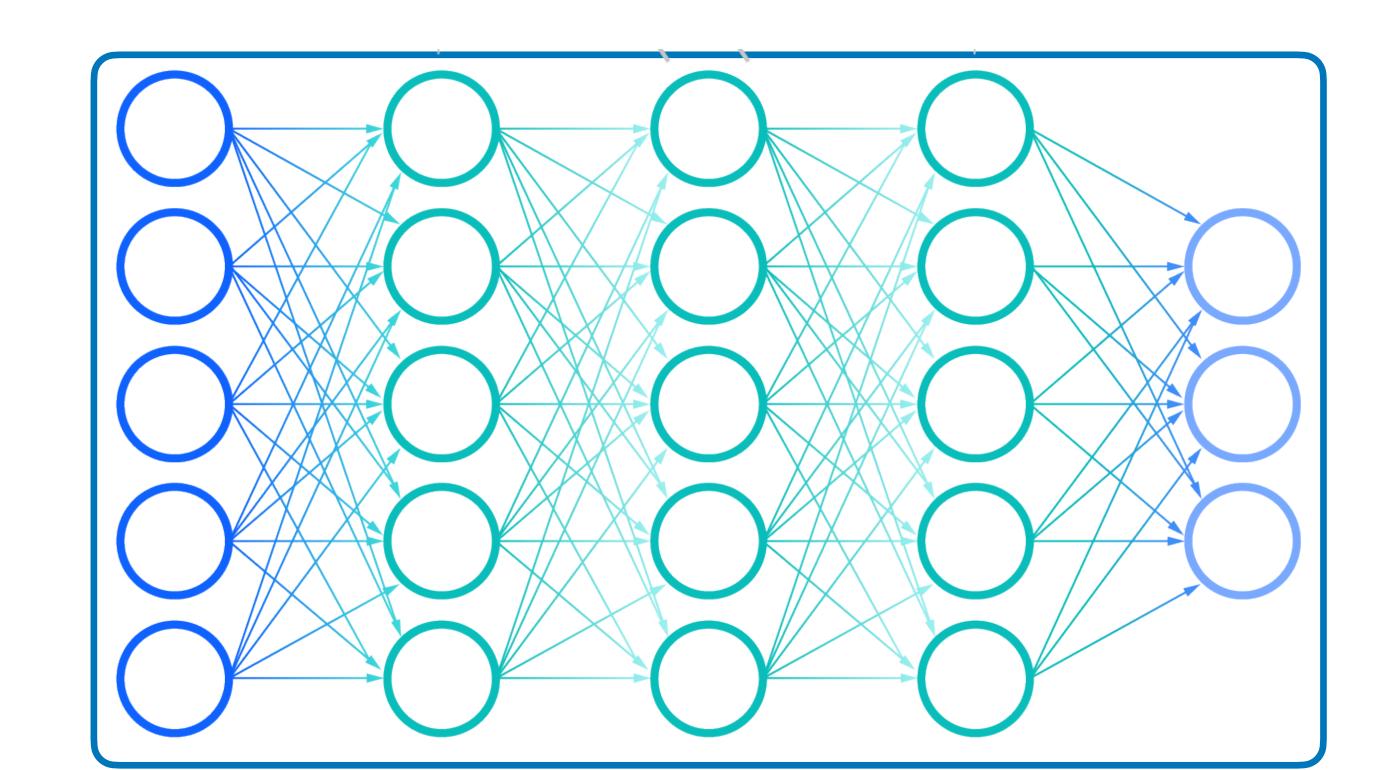
ımages

other species vocalizations (human speech, birds, other primates)

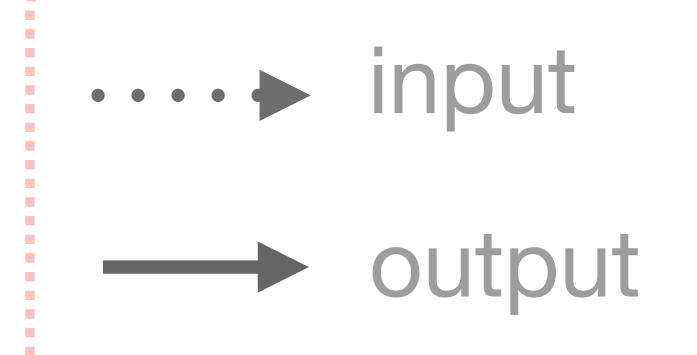
general sound

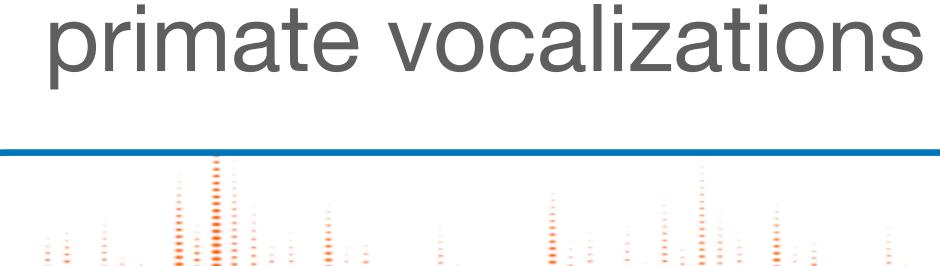
- VGG
- wav2vec Pre-training
- HuBERT
- BirdNet
- Vggish
- YAMNet
- AudioMAE

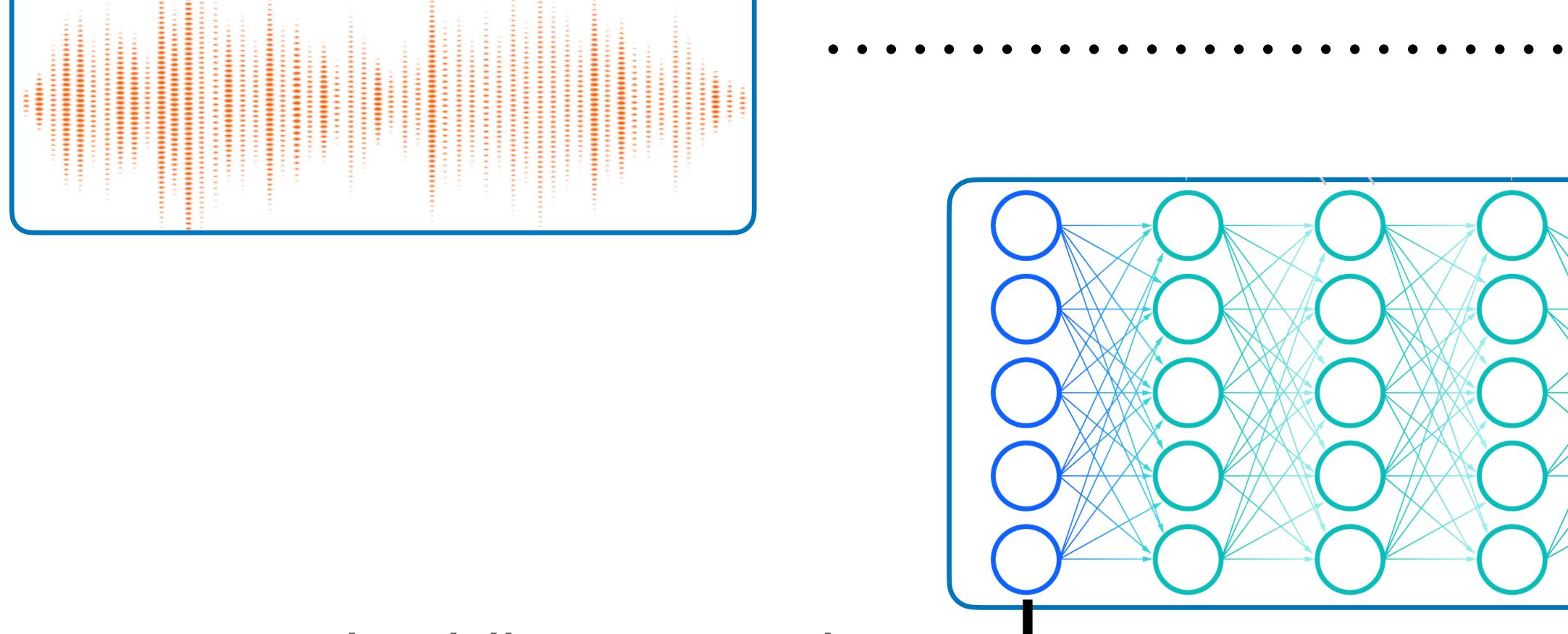
Pre-trained model



#2 Knowledge transfer

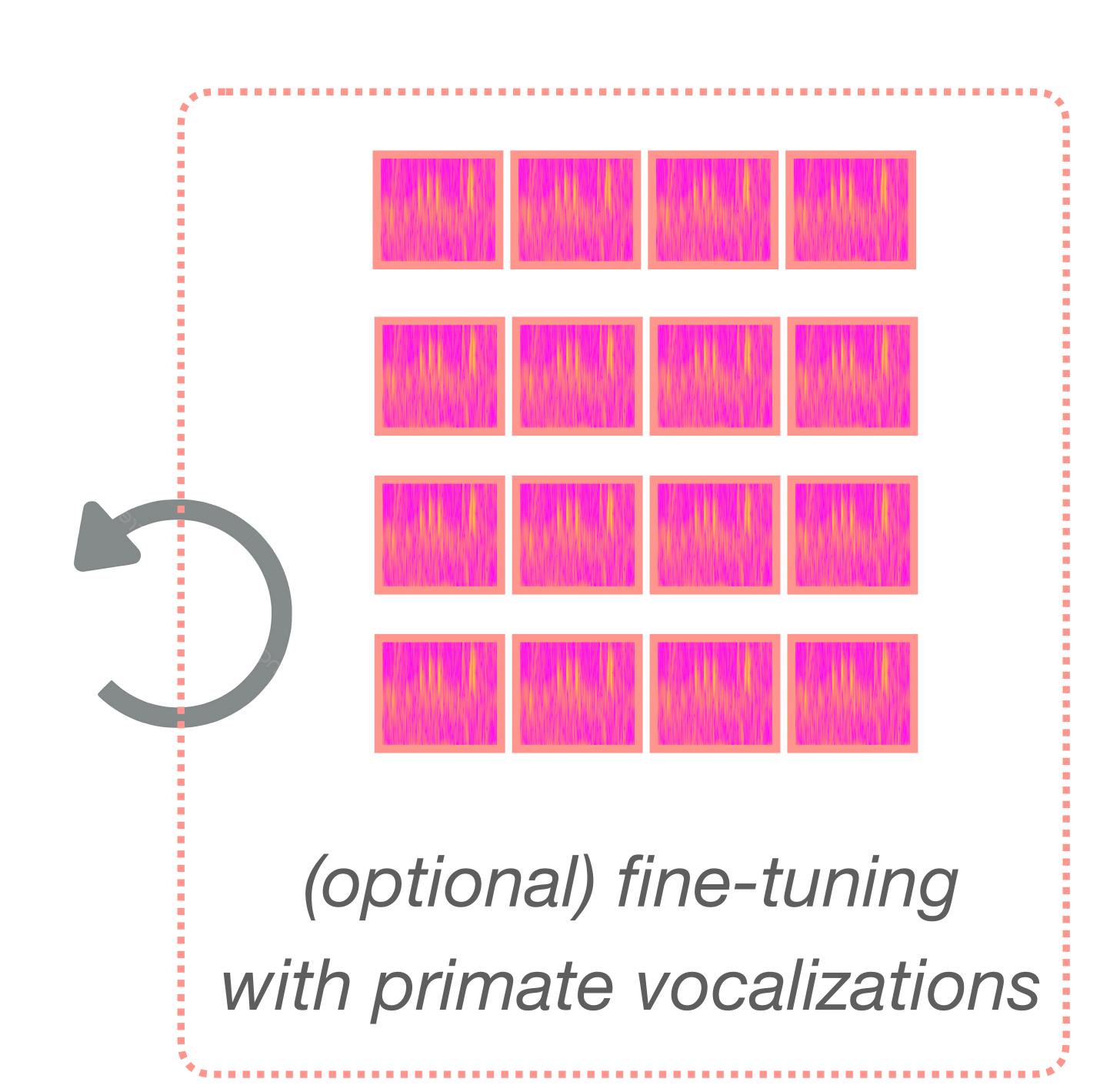




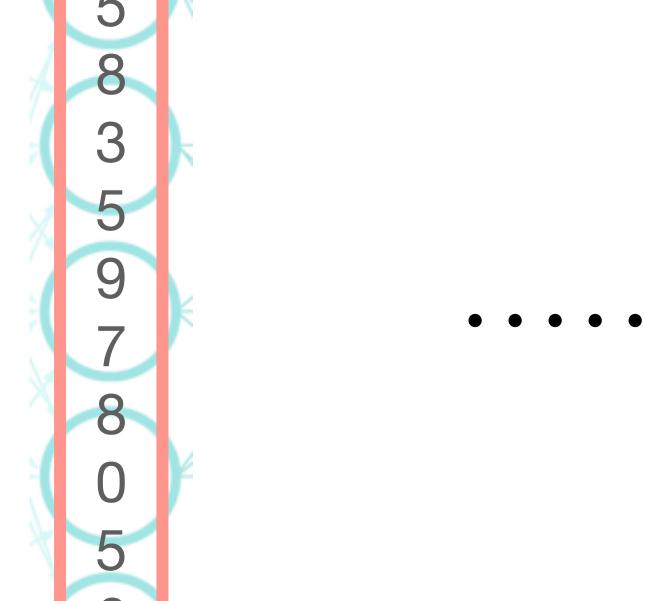


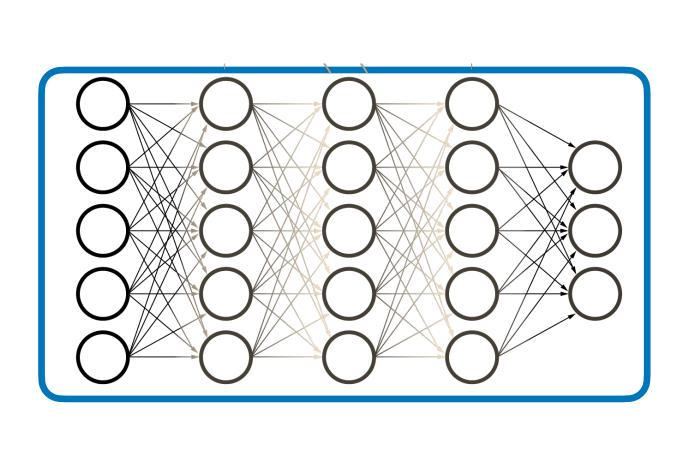
embedding extraction

pre-trained model

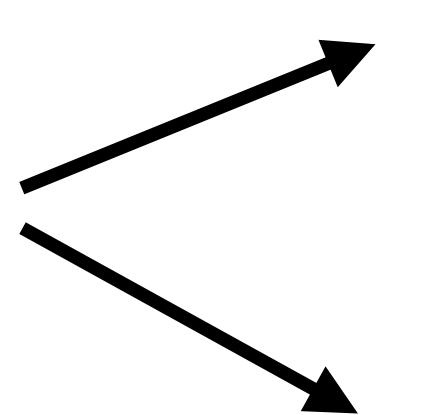


Label 1?

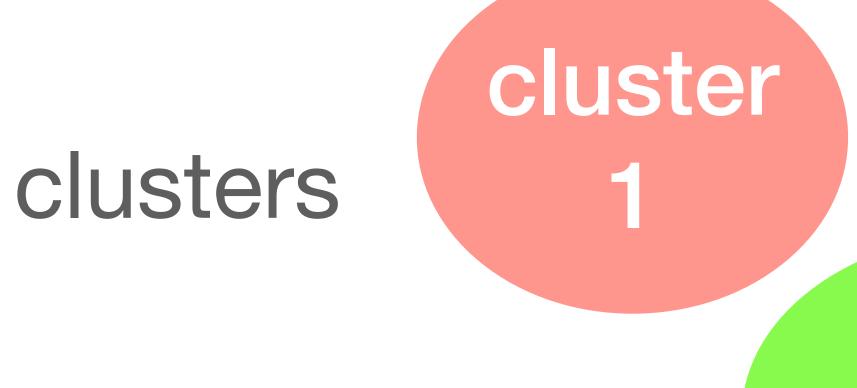




downstream model



predictions Label 2? Label 3?



cluster

cluster 2