



# *Frame-level Instrument Recognition by Timbre and Pitch*

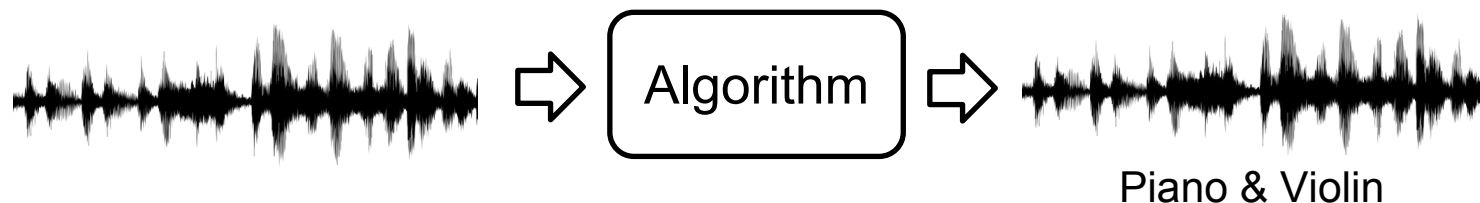
Yun-Ning (Amy) Hung  
<http://mac.citi.sinica.edu.tw/>  
[biboamy@iis.sinica.edu.tw](mailto:biboamy@iis.sinica.edu.tw)

**Music & Audio Computing Lab,**  
Research Center for IT Innovation,  
Academia Sinica

# Goal – Frame-level Instrument Recognition

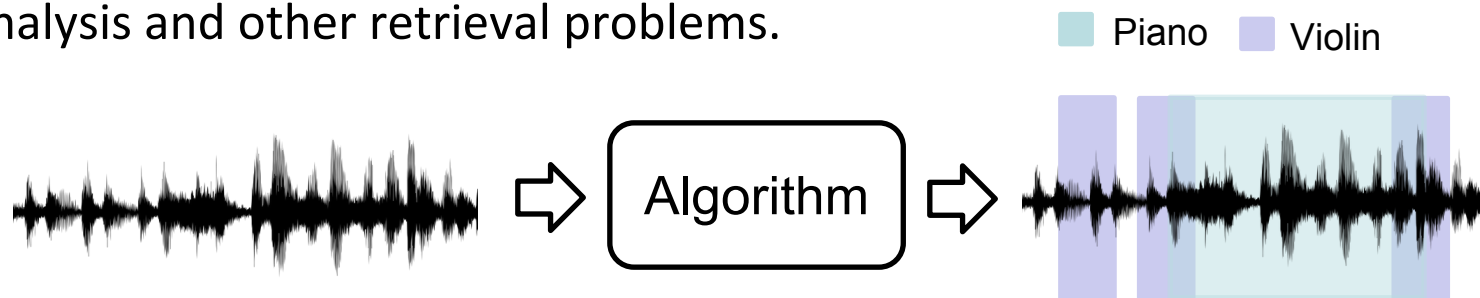
- Previous methods

Given an audio, algorithm will tell you which instruments are in the audio



- Our methods

Given an audio, algorithm will tell you the instruments start and end at which time frame. It can benefit music transcription, music structure analysis and other retrieval problems.



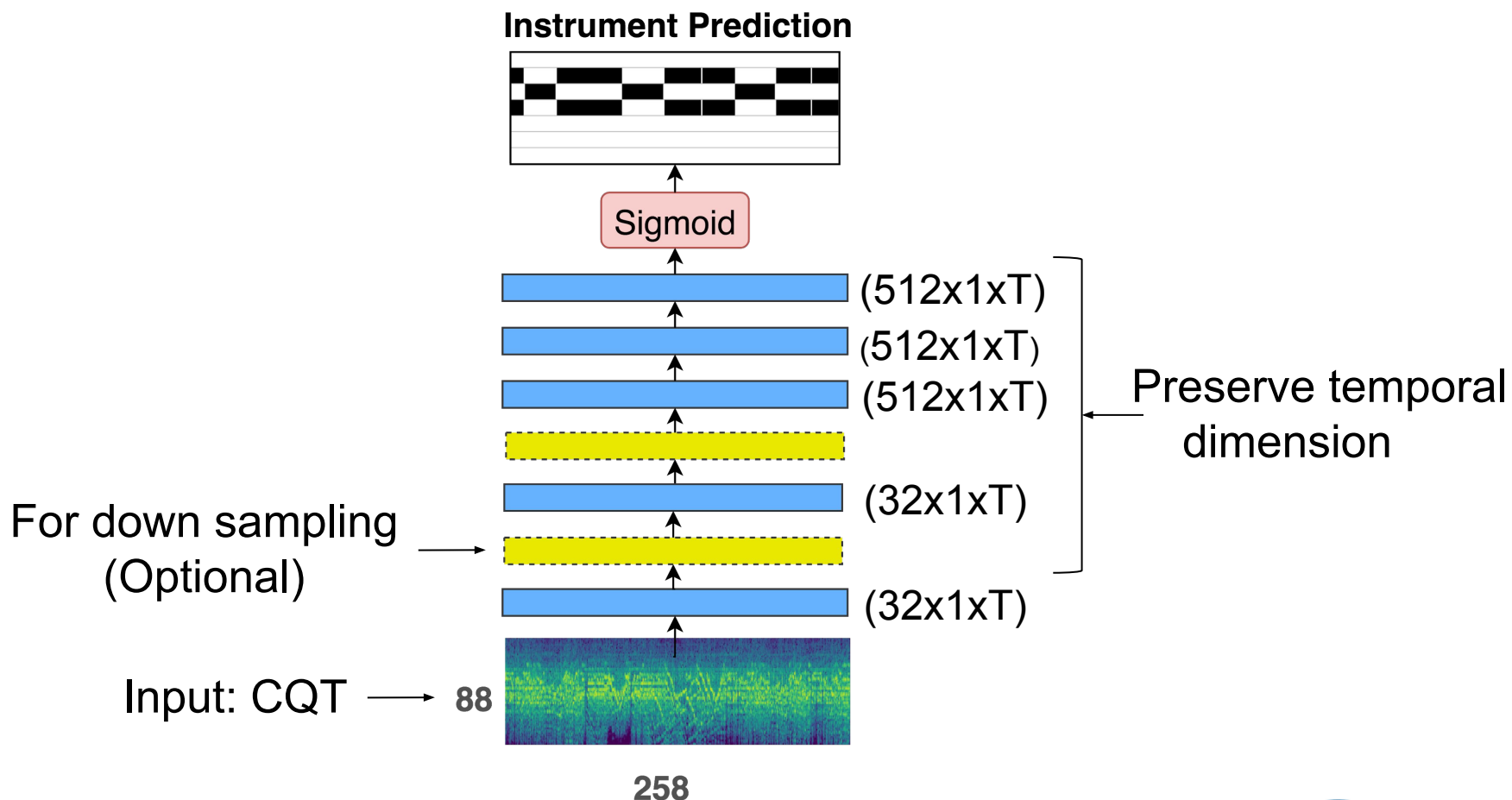
# Dataset

Dataset	Multi instrument	Instrument Frame Labels	Pitch Frame Labels
ParisTech, UIOWA, RWC			
IRMAS, AudioSet, MagnaTagATune	✓		
MedleyDB	✓	✓	Part of it
Mixing Secret	✓	✓	
MusicNet	✓	✓	✓

- Most instrument related dataset don't have frame labels
- We choose MusicNet dataset for two reasons:
  - Three times larger than other datasets which has frame labels
  - Has pitch labels



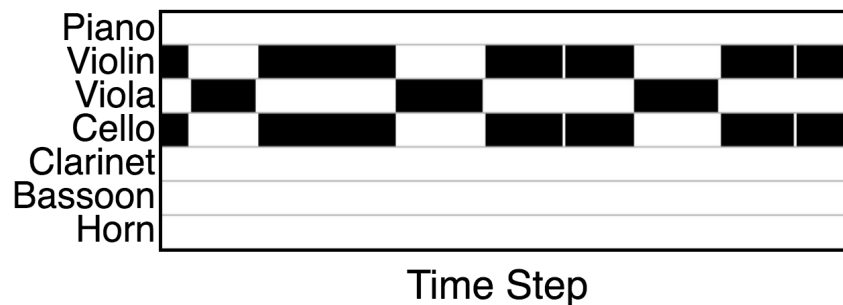
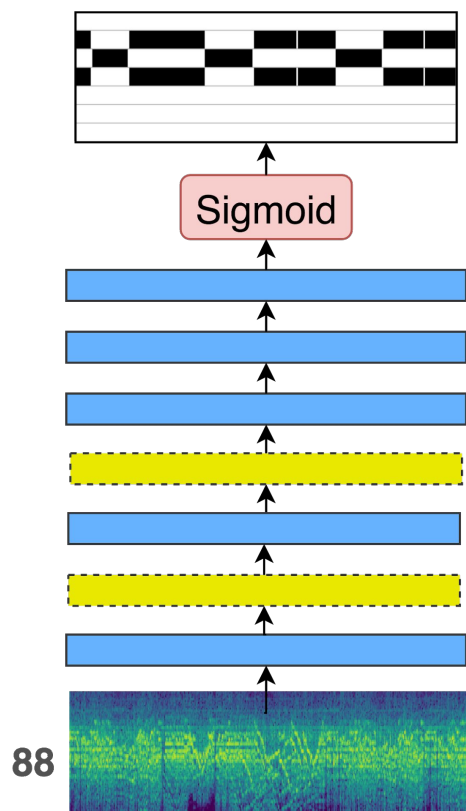
# Method – Convolutional Neural Network



# Method – Convolutional Neural Network

- Max Pool
- Conv + Batch Norm + ReLU

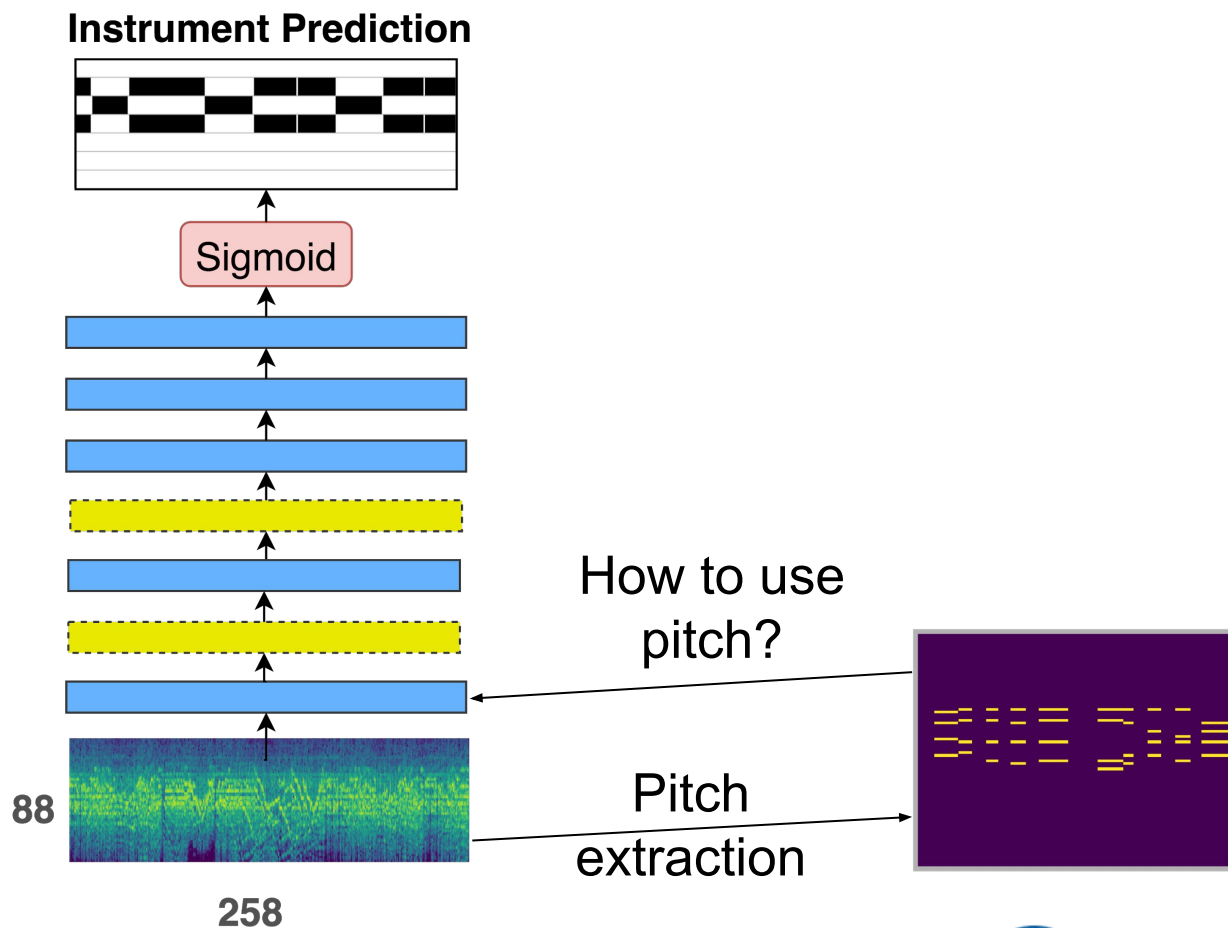
## Instrument Prediction



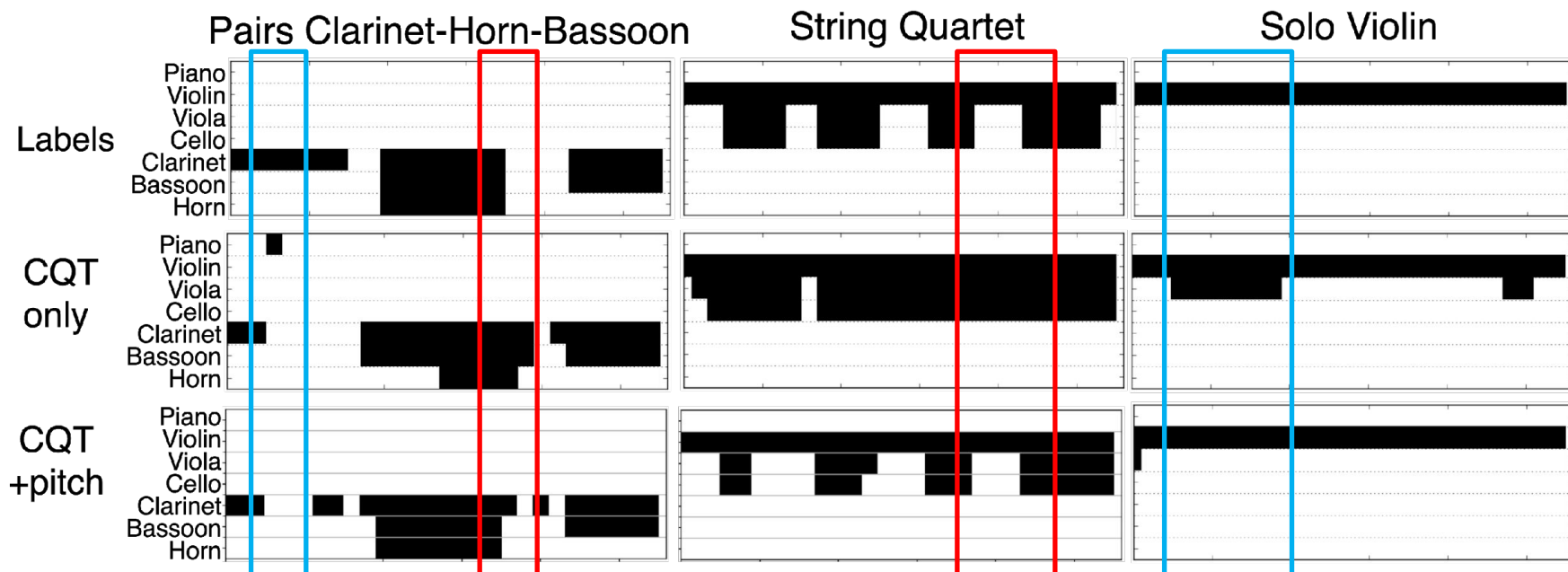
- Present of instrument in each time frame

# Method – Convolutional Neural Network

- Max Pool
- Conv + Batch Norm + ReLU



# Result – Output Tensor Representation



Help onset and offset detection

Ease the prediction error

# Poster: A-18

Thanks!

