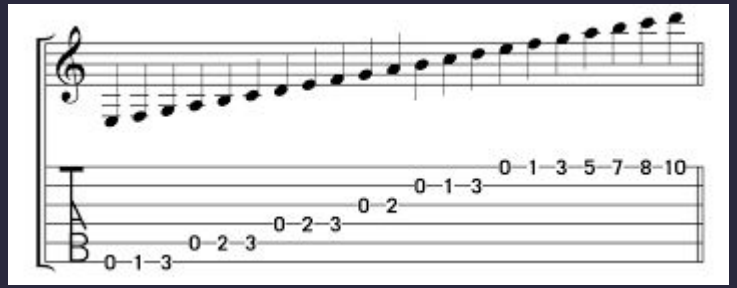
Automatic Guitar Transcription

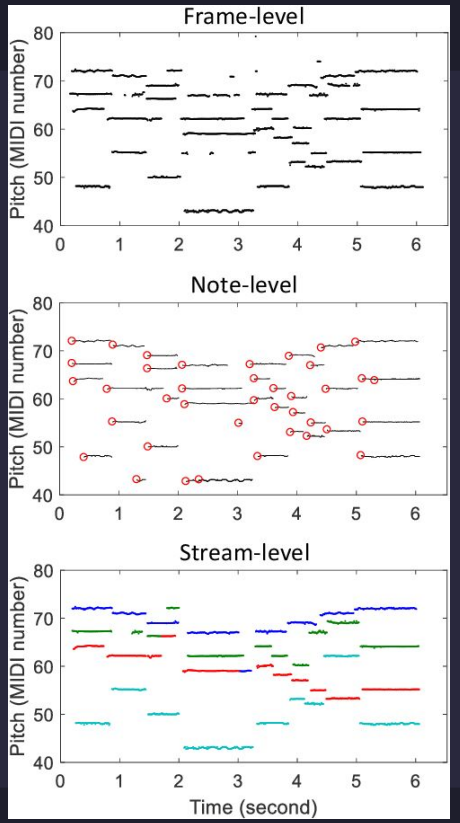
Seth Roberts

```
[[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
 [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]
```

Problem

- Automatic Music Transcription
- Abundance of piano data
- Tablature





Methods

Two Step Approach

1 - Pitch Estimation

2 - Converted to tab

One Step Approach

1 - Map audio directly to tab with CNN

Topics

01 GuitarSet

Dataset used for training

02 Compile Data

Get data from dataset and input to DataLoader

03 CNN

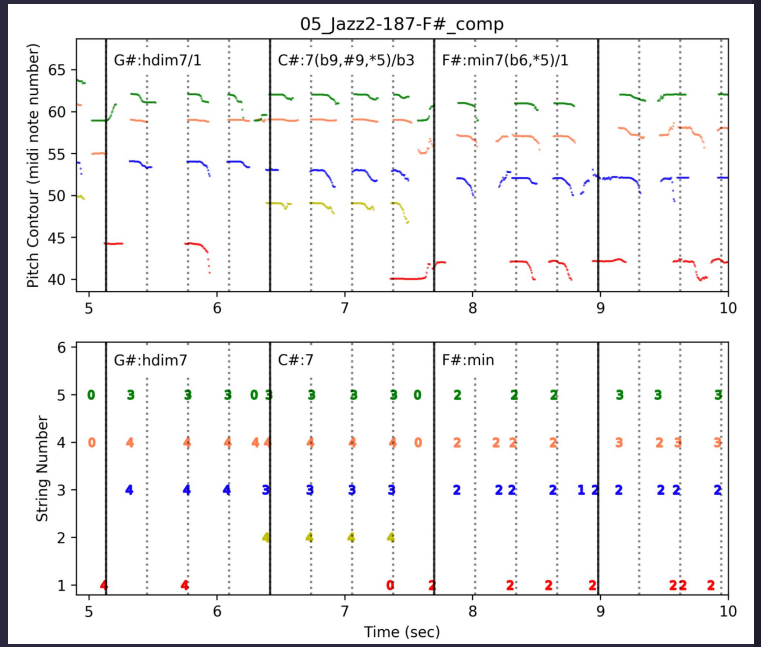
Convolutional Neural Network for transcription

04 Results

Analysis of output from model

GuitarSet

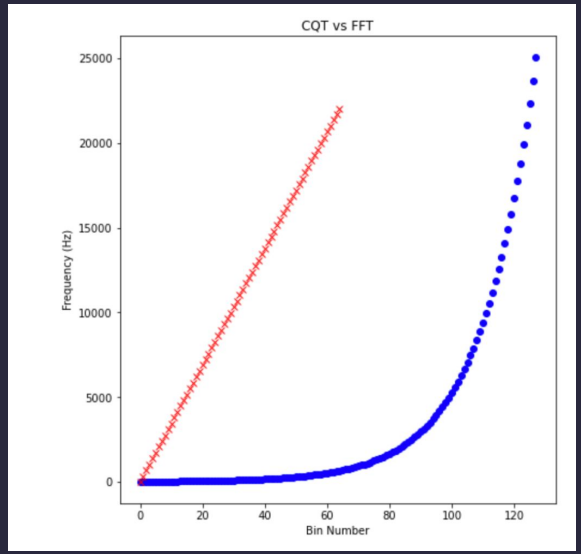
- 360 excerpts
- Various guitarists and genres
- Hexaphonic Pickup
- Annotation (Each string)
  - Midi Note
  - Beat Position



Data From GuitarSet

- **Audio Preprocessing**
  - Downsample
  - CQT
- **Labeling Preprocessing**
  - *.jams* files
  - 21 fret classes
  - 6 x 21 label array

[[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]

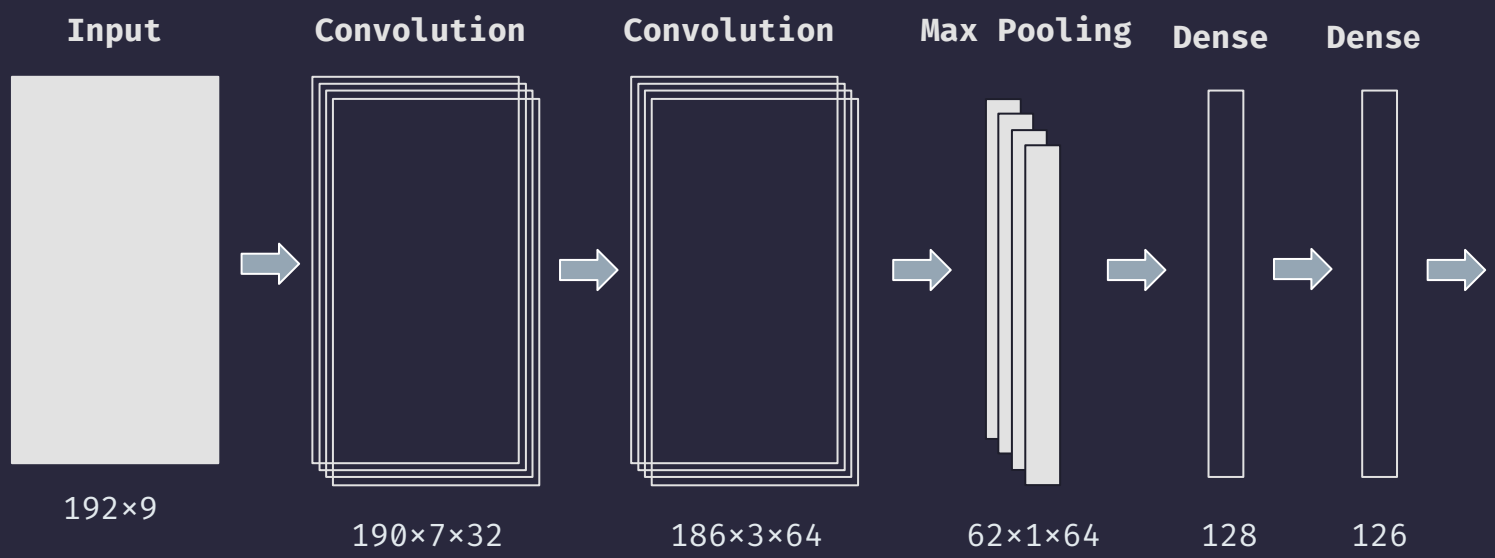


DataSet

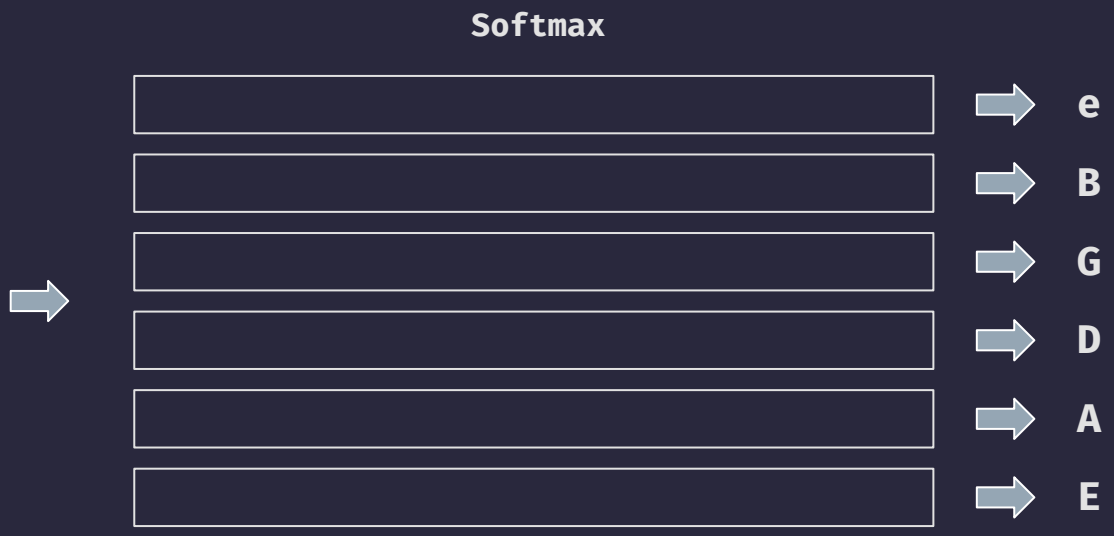
- Custom DataSet Class
- Context Window (192×9)
- List of Frames
- Train, Validate, Test

```
04_Jazz3-150-C_solo_0
04_Jazz3-150-C_solo_1
04_Jazz3-150-C_solo_2
04_Jazz3-150-C_solo_3
04_Jazz3-150-C_solo_4
04_Jazz3-150-C_solo_5
04_Jazz3-150-C_solo_6
```

Network Architecture



Network Architecture



wait....

Novelty

- Pytorch
- Different CNN Architecture
- Convolutional Layers
- Kernel Size
- Max Pooling
- Validation
- Optimization
- Dataset size

results ->

Results

- Dataset size significantly mattered
- Loss decreasing during training :)
- Loss increasing during validation :(
- Timeline of Project

Future

- Optimizer
- Learning Rate
- Batch Size
- Context Window Size
- Epochs

References

[1] Ana M Barbancho, Anssi Klapuri, Lorenzo J Tardon, and Isabel Barbancho. Automatic transcription of guitar chords and fingering from audio. IEEE Transactions on Audio, Speech, and Language Processing, 20(3):915–921, 2011.

[2]Chang, W., A. W. Su, C. Yeh, A. Roebel, and X. Rodet. 2008. Multiple-F0 tracking based on a high order HMM model. In Proceedings of the International Conference on Digital Audio Effects, Espoo, Finland.

[3]Yoonchang Han, Jaehun Kim, Kyogu Lee, Yoonchang Han, Jaehun Kim, and Kyogu Lee. Deep convolutional neural networks for predominant instrument recognition in polyphonic music. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 25(1):208–221, 2017.

[4]Eric J Humphrey and Juan P Bello. From music audio to chord tablature: Teaching deep convolutional networks to play guitar. In 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP), pages 6974–6978. IEEE, 2014.

[5]A. Klapuri, C. Schorkhuber, Constant-Q Transform Toolbox For Music Processing. University of London, 2010.

[6]J. Sleep, Automatic Music Transcription With COnvolutional Neural Networks Using Intuitive Filter Shapes, California Polytechnic State University, October 2017.

[7]Andrew Wiggins and Youngmoo Kim. Guitar tablature estimation with a convolutional neural network. In ISMIR, 2019.

[8]Q. Xi, R. Bittner, J. Pauwels, X. Ye, and J. P. Bello. Guitarset: A Dataset for Guitar Transcription. in 19th International Society for Music Information Retrieval Conference, Paris, France, Sept. 2018.

questions