A Data-Driven Methodology for Considering Feasibility and Pairwise Likelihood in Deep Learning Based Guitar Tablature Transcription Systems

Frank Cwitkowitz[1], Jonathan Driedger[2], and Zhiyao Duan[1]

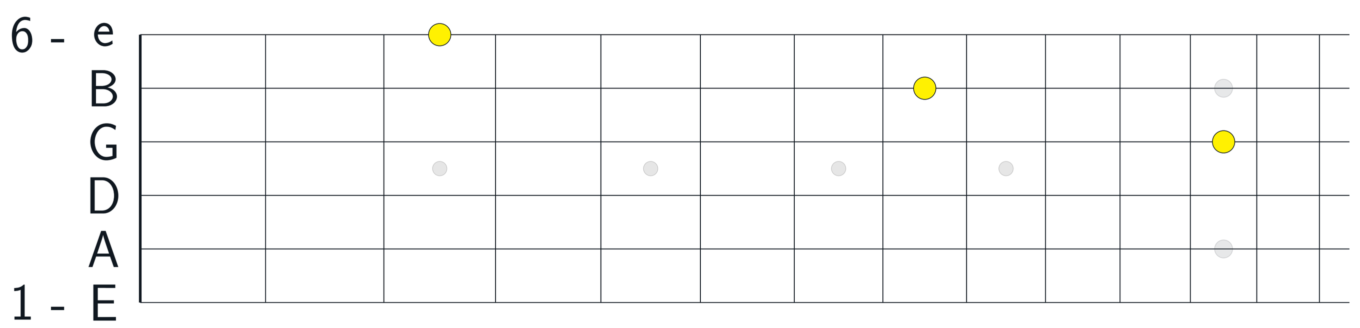
[1]Audio Information Research Lab, University of Rochester, [2]Chordify

Overview

- The pairwise likelihoods of string-fret (S/F) combinations are estimated using a large collection of symbolic tablature [1].
- A novel inhibition loss incorporating the estimated likelihoods is proposed for deep learning based models.
- The output layer of a baseline guitar tablature transcription model [2] is re-formulated and augmented with the inhibition loss.

Guitar Tablature Transcription

Generate a 6-hot vector $y_{s,f,n}$ for each frame n in a piece of audio, where 1s correspond to the chosen fret class $f \in \{-1, 0, 1, ..., F\}$ for each string $s \in \{1, ..., 6\}$. We use $c \in \{1, ..., C\}$ interchangeably to denote combinations of string and fret (S/F), where $C = 6 \times (F + 2)$.



Output Layer Formulation

Contemporary tablature transcription models [2, 3] apply the **softmax activation** across fret classes for each string at the output layer.

$$L_{CCE} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{s=1}^{6} \log\left(z_{s,f',n}\right) \tag{1}$$

This treats transcription as 6 independent classification tasks, ignoring the typically high correlation between the S/F combinations making up a fingering. We re-formulate the output layer using **sigmoid activations**, allowing us to introduce a novel inhibition loss.

$$L_{BCE} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{C} t_{c,n} \log\left(z_{c,n}\right) + (1 - t_{c,n}) \log\left(1 - z_{c,n}\right) \tag{2}$$

Datasets

DadaGP [1]

- Large collection of *GuitarPro* files featuring tablature for many popular full-length songs.
- Includes artists spanning many musical styles, with a bias toward rock and metal.
- We process all guitar tracks in standard tuning, yielding 33967 pieces of symbolic tablature.

GuitarSet [4]

- Comprises roughly 3 hours of acoustic guitar audio with string-level note annotations.
- Features 6 guitarists playing 2 unique interpretations over 30 different chord progressions.

Estimating Pairwise Likelihood

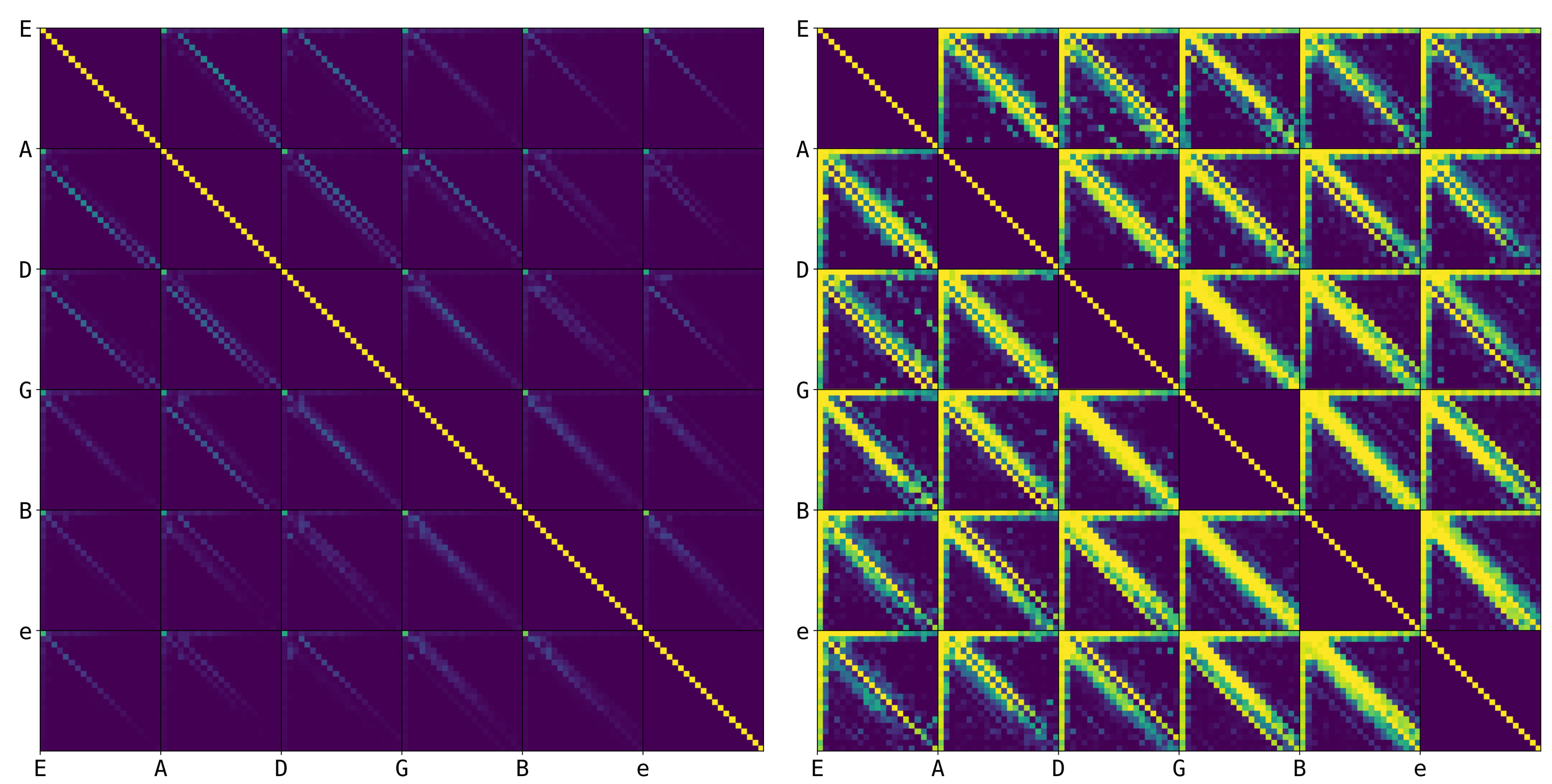
We can estimate the pairwise likelihood of two S/F combinations c_i and c_j using an arbitrary collection of symbolic tablature data (*e.g.*, DadaGP [1]). Given the symbolic tablature for a single track, we compute the intersection over union (IoU) of frame-level occurrences for all pairs of S/F combinations.

$$inter(i, j) = \sum_{n=1}^{N} t_{c_i,n} \wedge t_{c_j,n} \quad union(i, j) = \sum_{n=1}^{N} t_{c_i,n} \vee t_{c_j,n} \tag{3}$$

Let $\mathcal{T}'(i, j)$ be the set of tracks where c_i and c_j, independently, each occur in at least one frame. The IoU of the pair is averaged across these valid tracks

$$IoU(i, j) = \frac{1}{|\mathcal{T}'(i, j)|} \sum_{t \in \mathcal{T}'(i,j)} \frac{inter(i, j)_t}{union(i, j)_t}, \tag{4}$$

where $|\mathcal{T}'(i, j)|$ is the cardinality of $\mathcal{T}'(i, j)$. Note this is only valid for pairs where $|\mathcal{T}'(i, j)| > 0$. We set $IoU(i, j) = 0$ for all pairs where $|\mathcal{T}'(i, j)| = 0$.



Inhibition Loss

We introduce a novel loss term to inhibit the co-activation of unlikely pairs:

$$L_{inh} = \frac{1}{2N} \sum_{n=1}^{N} \sum_{i=1}^{C} \sum_{j=1}^{C} z_{c_i,n} z_{c_j,n} w(c_i, c_j). \tag{5}$$

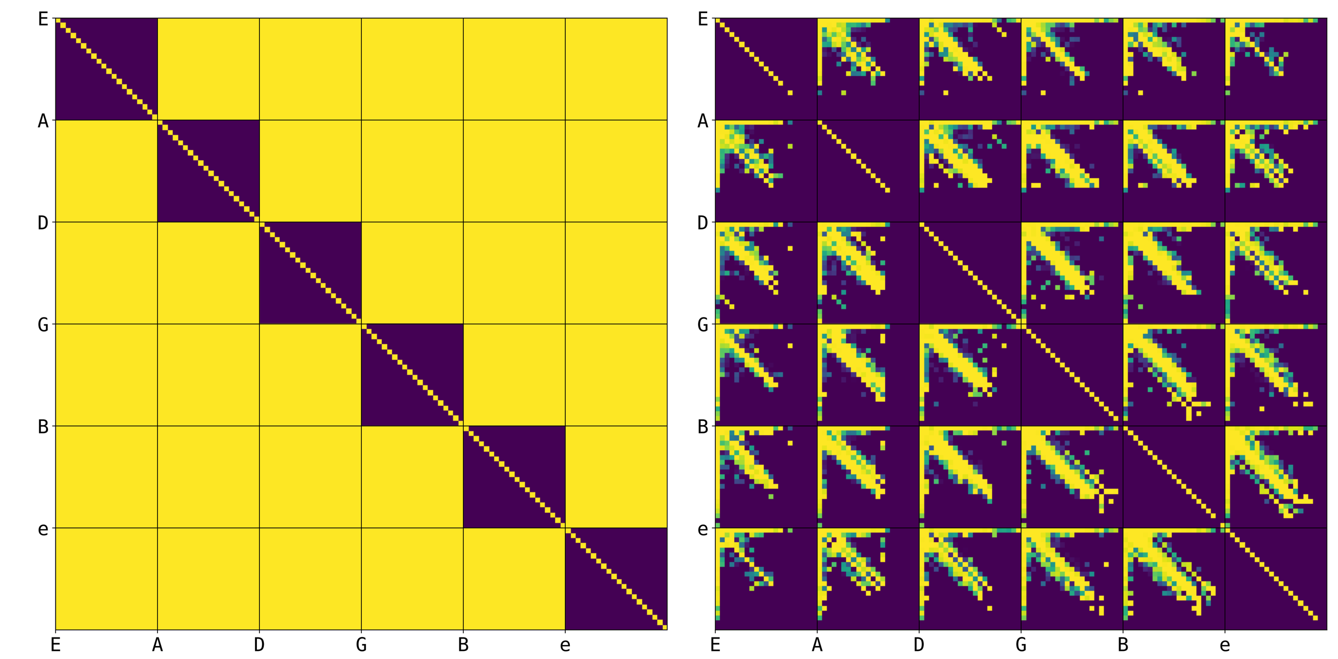
The product for every combination of activations is taken and scaled by an inhibition weight, a penalty between 0 and 1 for producing high activations for the combinations in the pair in a single frame. The result is summed over all combinations. We set the inhibition weights to be the complement of the pairwise likelihoods estimated using Equation (4), boosted with parameter b:

$$w(c_i, c_j) = (1 - IoU(i, j))^b. \tag{6}$$

Including a scaling term λ for balancing the two terms, the total loss becomes

$$L_{total} = L_{BCE} + \lambda L_{inh}. \tag{7}$$

Experiments



- Employ TabCNN [2] as a baseline model for guitar tablature transcription.
- Train and evaluate on GuitarSet [4] following 6-fold cross-validation schema [2].
- Experiment with holding out an extra dataset split for validation.
- Experiment with inserting a uni-directional LSTM before the output layer.
- Experiment with variations of the proposed output layer formulation.
- Adopt the metrics proposed in [2], but average across tracks, then folds.
- Compute inhibition losses L_{inh} $(b = 1)$ and $L_{inh}{}^{+}$ $(b = 2^7)$ on final predictions.
- Count number of duplicate pitch $E_{d.p.}$ and false alarm $E_{f.a.}$ errors.

Tablature Output Layer	p_{tab}	r_{tab}	f_{tab}	p_{pitch}	r_{pitch}	f_{pitch}	TDR	L_{inh}	$L_{inh}{}^{+}$	$E_{d.p.}$	$E_{f.a.}$
Softmax	0.809	0.692	0.742	0.910	0.762	0.825	0.903	8.87	0.132	21.4	359.8
Softmax w/ Val.	0.775	0.696	0.730	**0.895**	0.781	0.830	0.886	**9.01**	0.152	34.2	**442.5**
Softmax w/ Val./Rec.	0.783	0.757	0.768	0.879	0.835	0.854	0.905	9.27	0.158	24.3	489.6
Sigmoid $(\lambda = 0)$	0.782	0.757	0.767	0.878	0.836	0.854	0.902	9.27	0.154	20.0	503.3
Sigmoid w/ S $(\lambda = 1)$	**0.789**	**0.761**	**0.773**	0.881	**0.836**	**0.856**	**0.907**	9.25	0.155	19.5	485.8
Sigmoid w/ D $(\lambda = 1)$	0.787	0.743	0.763	0.880	0.821	0.847	0.902	9.19	0.147	12.0	481.8
Sigmoid w/ D^{+} $(\lambda = 1)$	0.782	0.754	0.766	0.876	0.833	0.852	0.902	9.25	0.143	13.8	496.6
Sigmoid w/ D^{+} $(\lambda = 10)$	0.781	0.755	0.766	0.867	0.829	0.845	0.907	9.26	**0.132**	**10.6**	504.6

The lack of a solid increase in tablature performance when using the inhibition objective can most likely be attributed to the small size of GuitarSet [4], the presence of some noisy labels which include duplicate pitch errors, and the difference between the distribution of DadaGP [1] and GuitarSet [4]. Overall, we argue that the lower $E_{d.p.}$ and $L_{inh}{}^{+}$ suggests that models trained with D and D^{+} produce tablature which is more feasible to play and more consistent with DadaGP [1].

Acknowledgements & References

This work is partially funded by National Science Foundation grants IIS-1846184 and DGE-1922591.
All of the code is available at `https://github.com/cwitkowitz/guitar-transcription-with-inhibition`.

[1] Pedro Sarmento et al. "DadaGP: A Dataset of Tokenized GuitarPro Songs for Sequence Models". In: *Proceedings of ISMIR*. 2021.

[2] Andrew Wiggins and Youngmoo Kim. "Guitar Tablature Estimation with a Convolutional Neural Network". In: *Proceedings of ISMIR*. 2019.

[3] Eric J Humphrey and Juan P Bello. "From Music Audio to Chord Tablature: Teaching Deep Convolutional Networks to Play Guitar". In: *Proceedings of ICASSP*. 2014.

[4] Qingyang Xi et al. "GuitarSet: A Dataset for Guitar Transcription". In: *Proceedings of ISMIR*. 2018.