



Deep Learning for Symbolic Music Representations

Néstor Nápoles López, PhD



About me



Submitted my PhD dissertation on August 15, 2022 (McGill University)



Focus on computational music theory



Started to educate myself about deep learning since 2019



Currently working in the industry (Avid Technology – Sibelius)

Outline

- Introduction to symbolic music data
 - Representations
 - Symbolic vs Audio Data
 - Symbolic music tasks
- Concrete example
 - Automatic Roman Numeral Analysis
 - End-user applications
- Additional resources
 - Deep learning for music generation
 - Datasets
 - Libraries

Outline

- **Introduction to symbolic music data**
 - Representations
 - Symbolic vs Audio Data
 - Symbolic music tasks
- Concrete example
 - Automatic Roman Numeral Analysis
 - End-user applications
- Additional resources
 - Deep learning for music generation
 - Datasets
 - Libraries

What is a “symbolic music file”?

A file that encodes musical information in a machine-readable representation

Symbolic music format	Applications
ABC	Short and concise for single melodies
Humdrum (**kern)	Great for data science
Lilypond	LaTeX-like markup
MIDI	Ubiquitous in music hardware/performance
MEI	Web workflows / Digital editions
MusicXML	Exchange format for music notation software
Proprietary (.sib, .gp, .mus, .mscz)	Application specific formats

Examples

- Folk RNN
 - <https://folkrnn.org/tune/>
 - Produces **ABC** folk tunes in your browser
- Verovio Humdrum Viewer
 - <https://verovio.humdrum.org/>
 - A visualizer of Humdrum/MEI/MusicXML data
- Hacklily:
 - <https://www.hacklily.org/>
 - Web engraver of Lilypond data

Symbolic and Audio

Symbolic data

- Music score or music performance
- Musical information readily available
- Metadata (composer, title, etc.)

Audio

- Waveform
- Air-pressure over time
- Hard to retrieve any musical information

Scope of symbolic music research

- Usually more in the terrain of musicologists and music theorists than engineers
- More granular questions than the audio domain equivalents



Domains in contrast

Audio

- “What is the chord we hear at this timestep?”

Symbolic

- “What is the function of the chord in the current musical passage?”



Domains in contrast

Audio

- “Extract the melody from this flute solo”

Symbolic

- “Extract the motivic patterns from this flute solo”



Domains in contrast

Audio

- “Find the musical key in this track”

Symbolic

- “Find the changes of musical key (modulations) in this track”

Symbolic music tasks

- Chorale generation
 - [BachBot](#)
 - [CoCoNet](#)
 - [DeepBach](#)
- Conditional music generation
 - [Music Transformer](#)
 - [Anticipatory Music Transformer](#)
- Expressive performance (predict dynamics given the notes)
- Pattern identification/clustering
- Score following
- Piano fingering prediction
- ...

Outline

- Introduction to symbolic music data
 - Representations
 - Symbolic vs Audio Data
 - Symbolic music tasks
- **Concrete example**
 - Automatic Roman Numeral Analysis
 - End-user applications
- Additional resources
 - Deep learning for music generation
 - Datasets
 - Libraries

Automatic Roman Numeral Analysis in Symbolic Music Representations

Summary

*Improvements to multitask learning methods for automatically
annotating a digital score with **Roman numeral analysis** labels*

Roman Numeral Analysis

- A form of harmonic analysis

Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists

Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists



Weber (1818, vol. 2, p. 37)

Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists

A musical score consisting of two staves of music. The top staff is in common time and has a key signature of one sharp. The bottom staff is also in common time and has a key signature of one sharp. Below the music, Roman numerals indicate the harmonic progression: C:I, IV: I, IV: II, I, V: V, I., a: I, IV, I. The Roman numerals C:I and a: I are highlighted with yellow boxes.

Weber (1818, vol. 2, p. 37)

Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists

A musical score consisting of two staves of music. The top staff shows a melody line with various note heads and stems. The bottom staff shows harmonic analysis using Roman numerals. The analysis is as follows:

C: I IV I IV_{II} I V₇ I. a:I IV I

The Roman numerals are highlighted in yellow.

Weber (1818, vol. 2, p. 37)

Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists

A musical score consisting of two staves of music. The top staff is in common time and has a key signature of one sharp. The bottom staff is also in common time and has a key signature of one sharp. Below the music, Roman numerals indicate harmonic progressions. The first measure is labeled 'I'. The second measure is labeled 'IV'. The third measure is labeled 'I'. The fourth measure is labeled 'V' with a yellow box around it. The fifth measure is labeled 'I.'. The sixth measure is labeled 'I'. The seventh measure is labeled 'IV'. The eighth measure is labeled 'I'.

Weber (1818, vol. 2, p. 37)

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

Largo

ff

5

p ritenuto.

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

Largo

ff

A♭ D♭

5

ritenuto.

9

p

A♭ D♭

23

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

Largo

Ab Db Eb⁷ Ab

Ab I IV V⁷ I

5

9

p ritenuto.

Ab Db

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

Largo

ff

A♭ Db E♭⁷ A♭

Ab I IV V⁷ I

5

9

p

ritenuto.

Ab Db G⁷ Cm

c:VI bII V⁷ i

Automatic Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists
- As a computational problem, it has been often described as
 - *key finding + chord labeling*

Automatic Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists
- As a computational problem, it has been often described as
 - *key finding + chord labeling + chord inversion + chord segmentation*

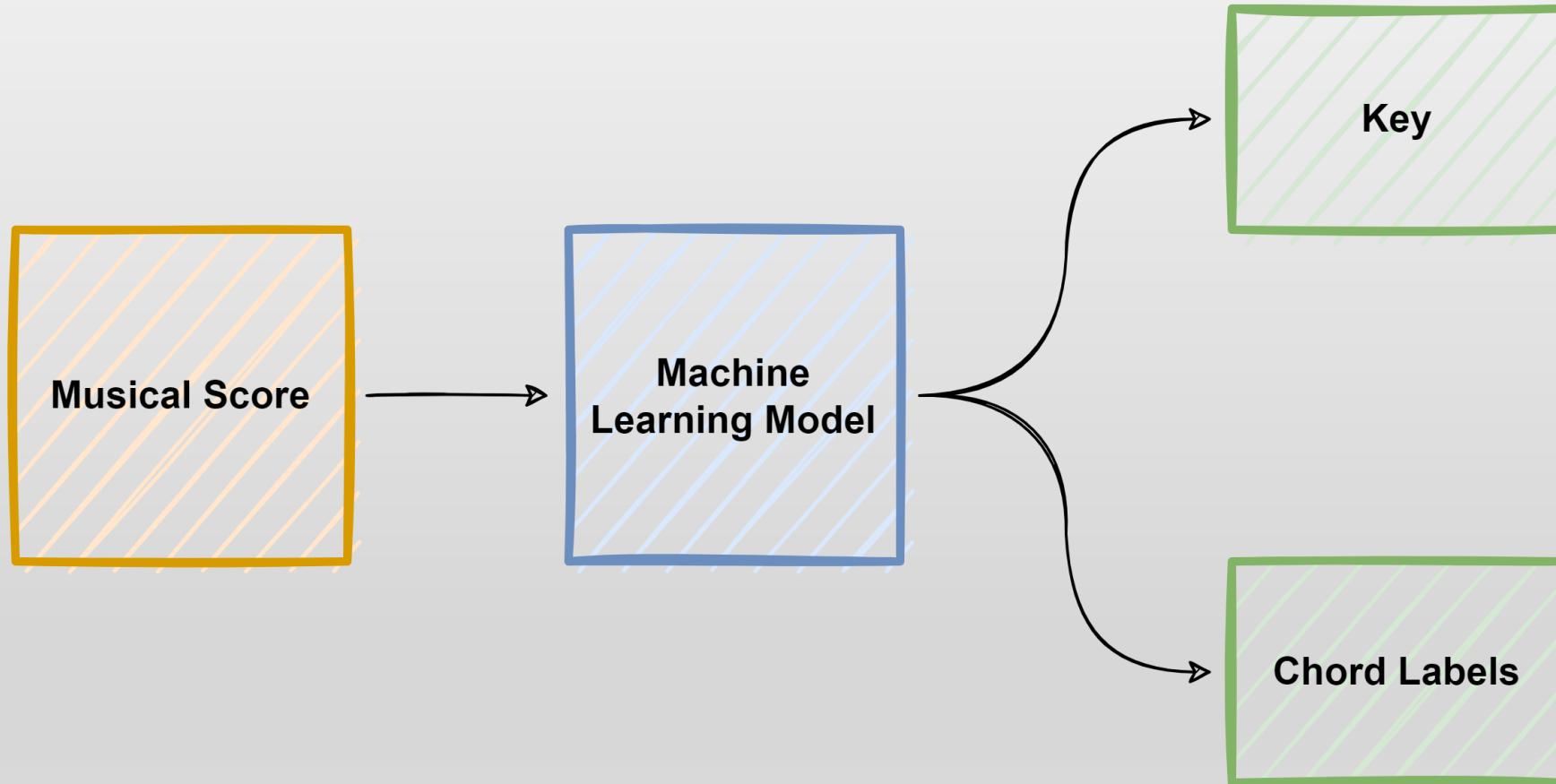
Automatic Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists
- As a computational problem, it has been often described as
 - *key finding + chord labeling*

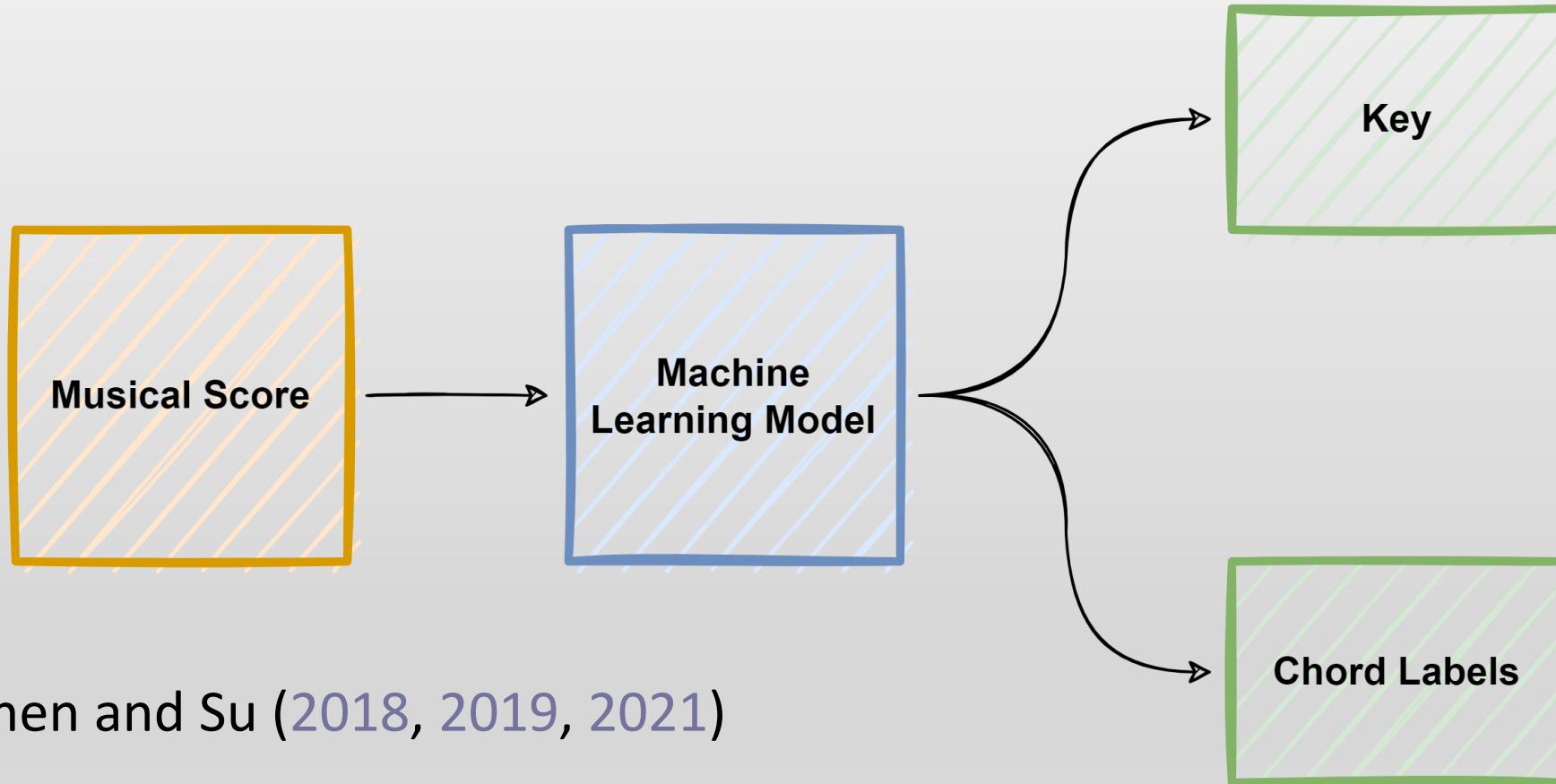
Automatic Roman Numeral Analysis

- A form of harmonic analysis
- The notation was introduced in the early 19th century and slowly adopted throughout the years by several music theorists
- As a computational problem, it has been often described as
 - *key finding + chord labeling*
- Using deep learning, it has been modeled through multitask learning CRNNs/Transformers or through modular classifiers (chord root, chord inversion, key, etc.)

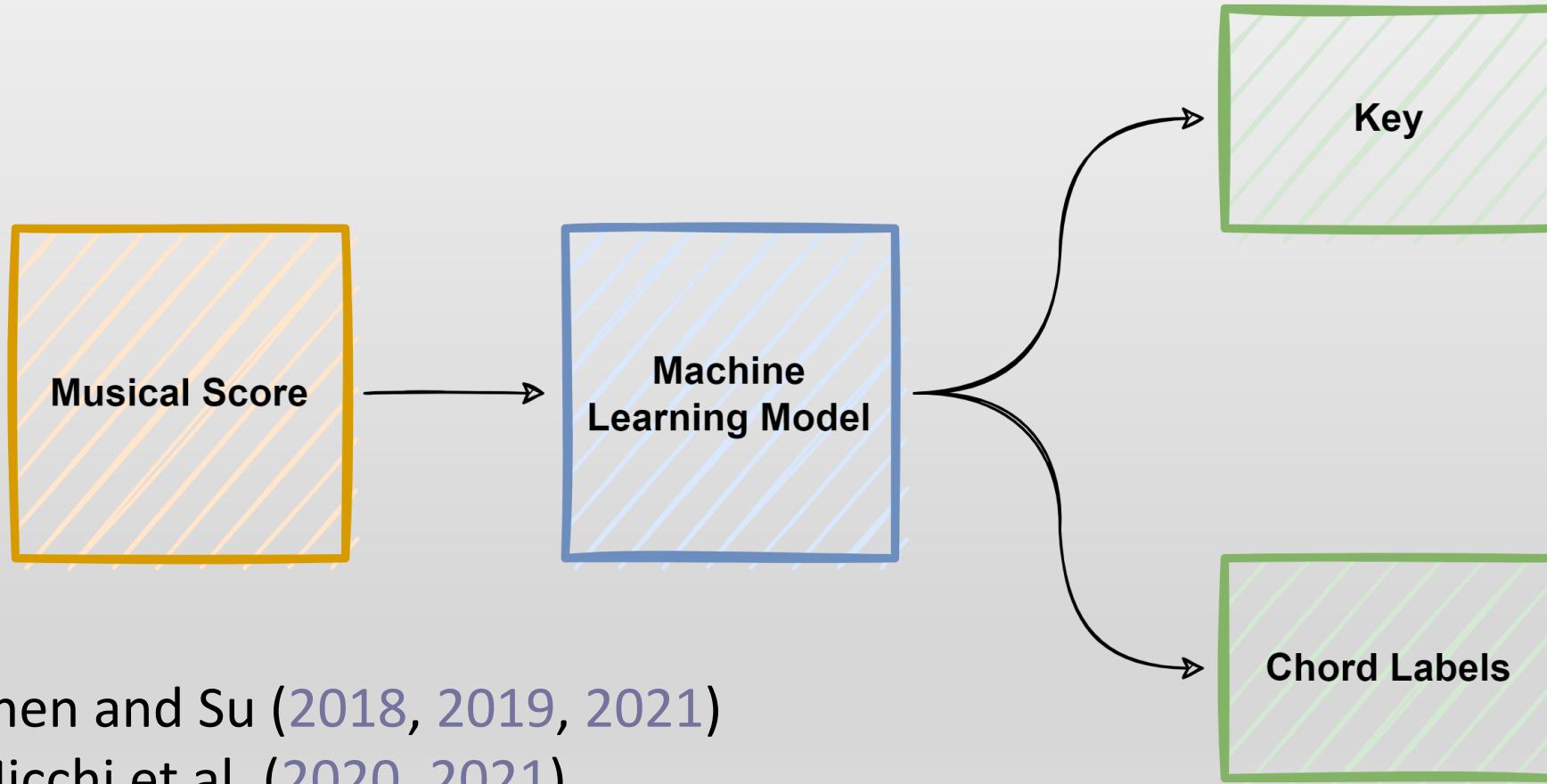
Multitask Learning Layout



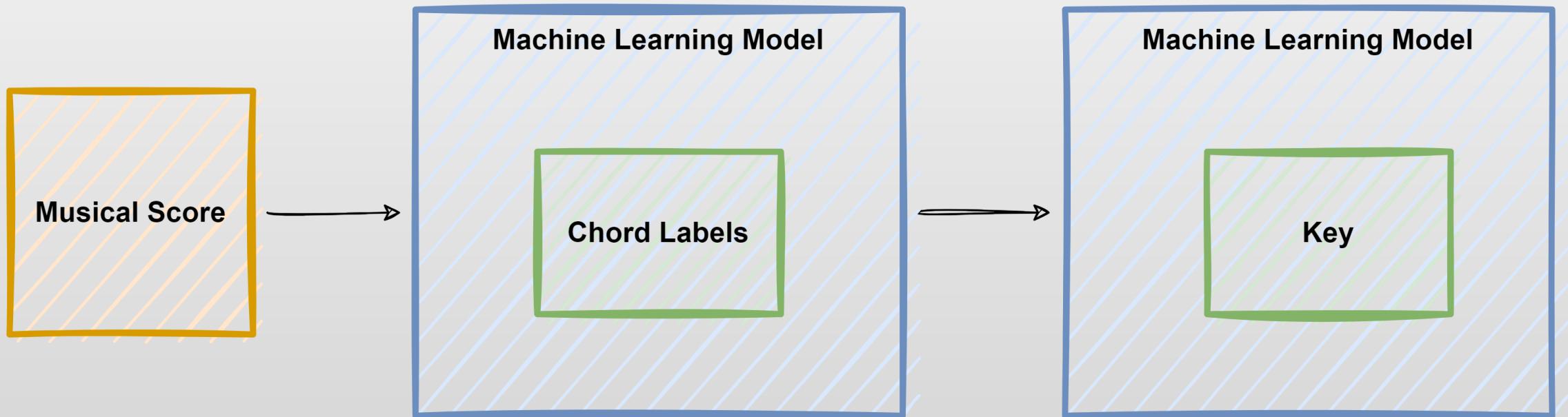
Multitask Learning Layout



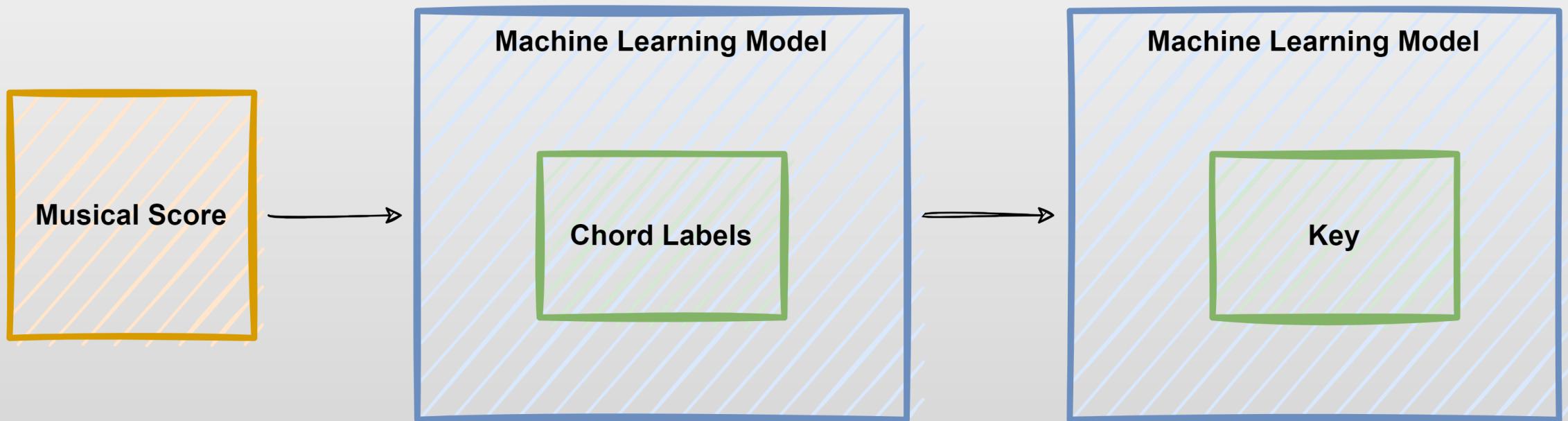
Multitask Learning Layout



Modular Analysis

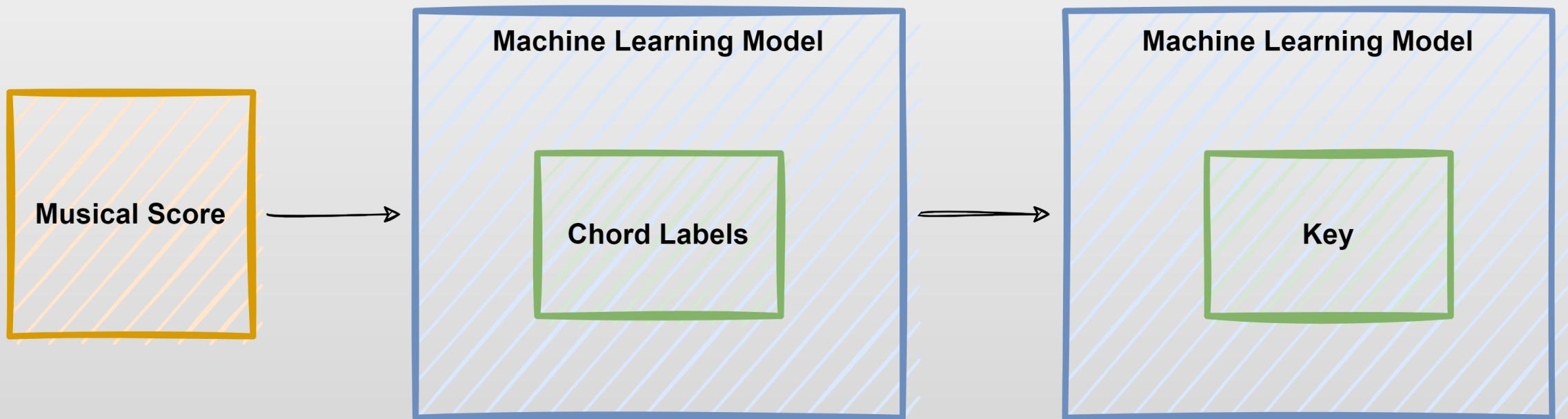


Modular Analysis



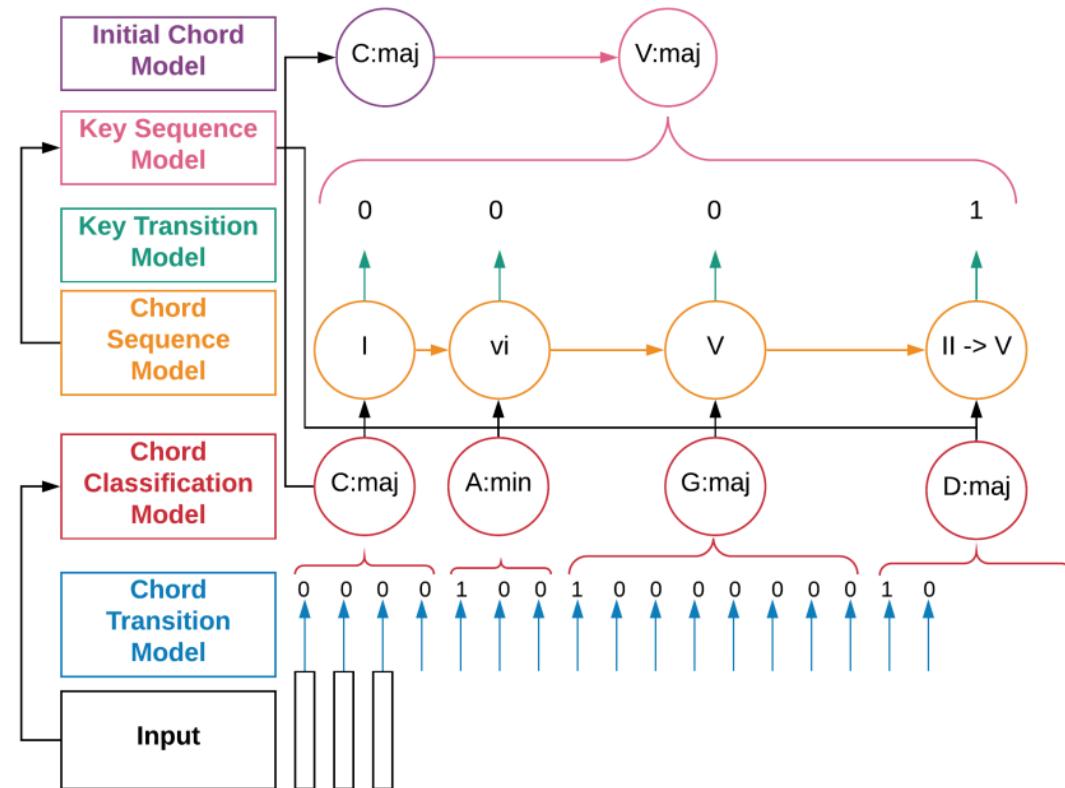
- Melisma (2003)

Modular Analysis

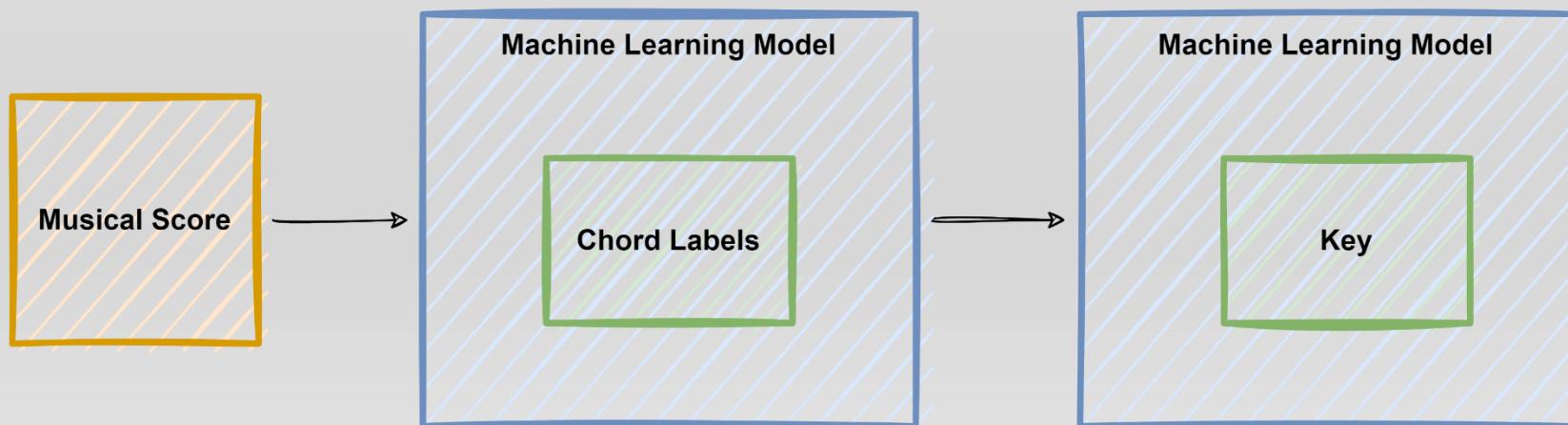
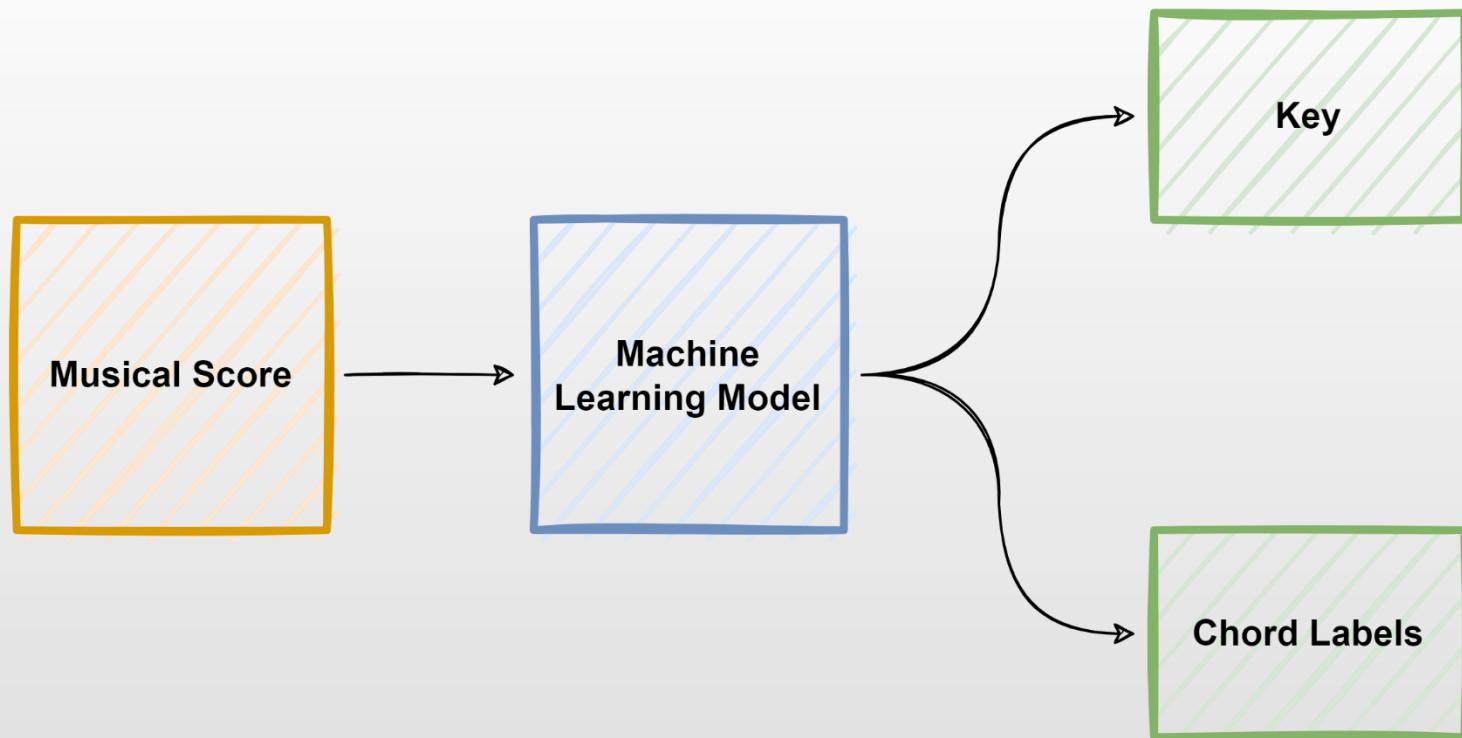


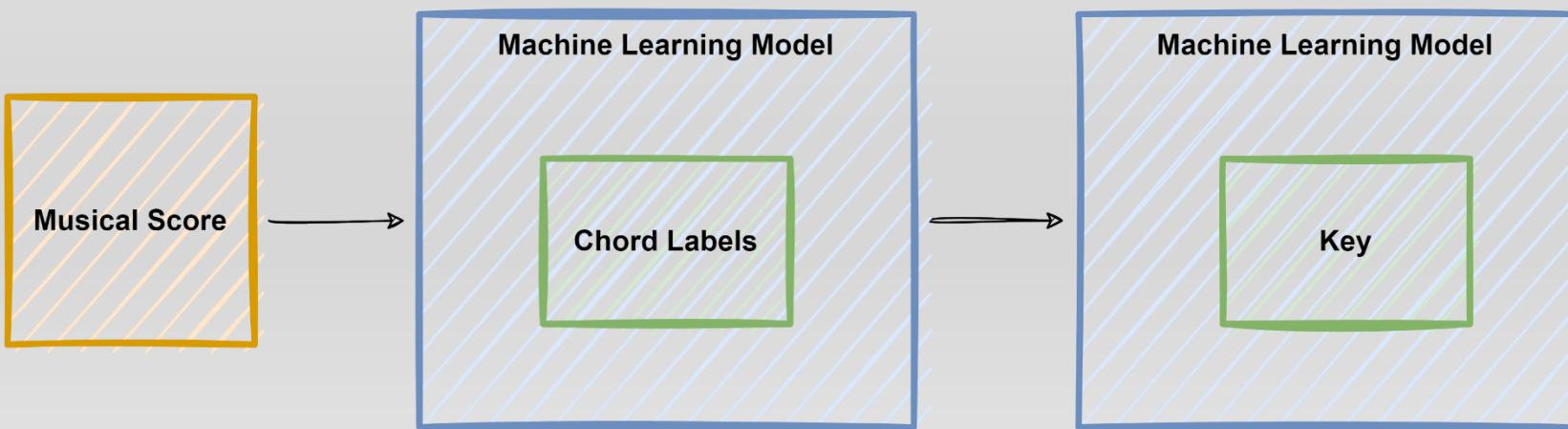
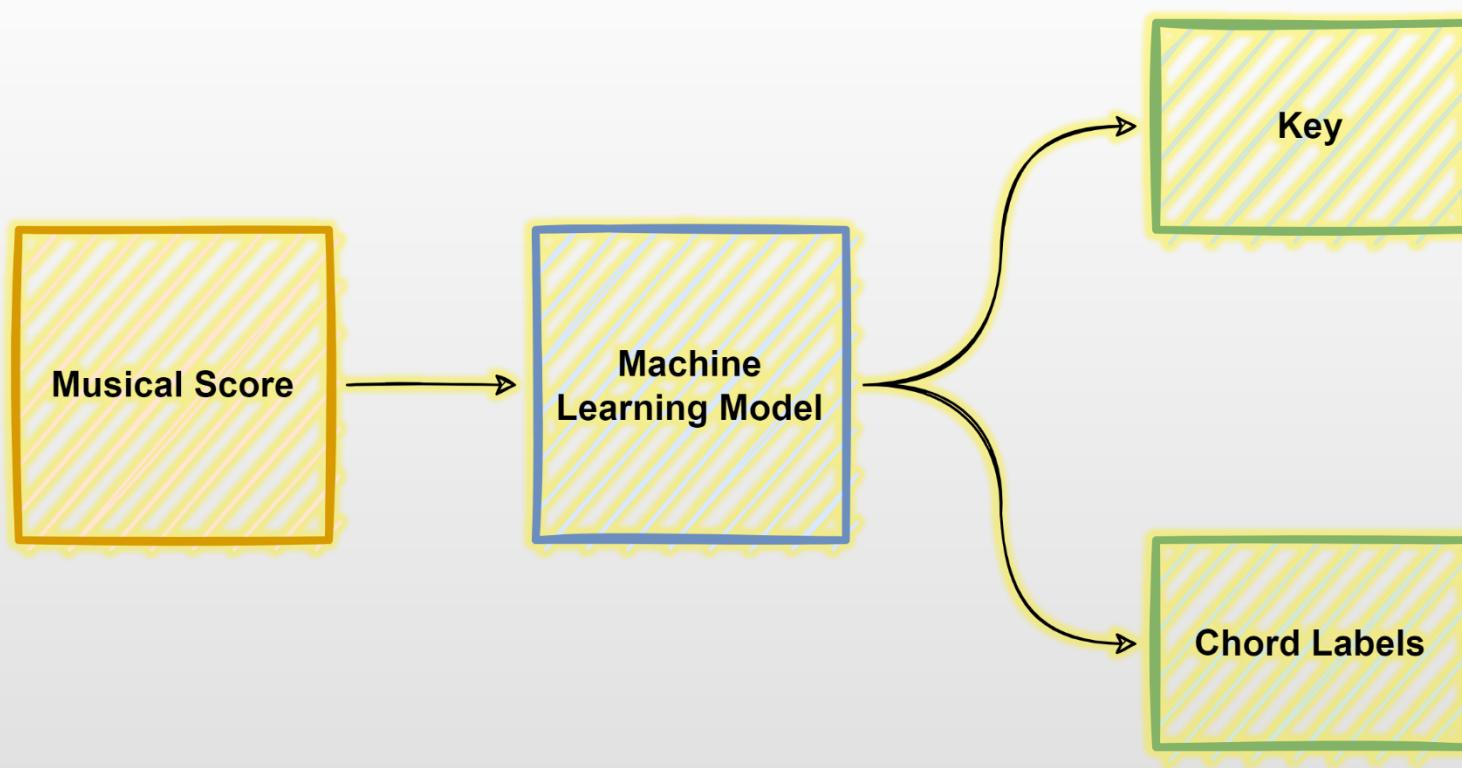
- Melisma (2003)
- McLeod and Rohrmeier (2021)

Modular Analysis



- McLeod and Rohrmeier (2021)

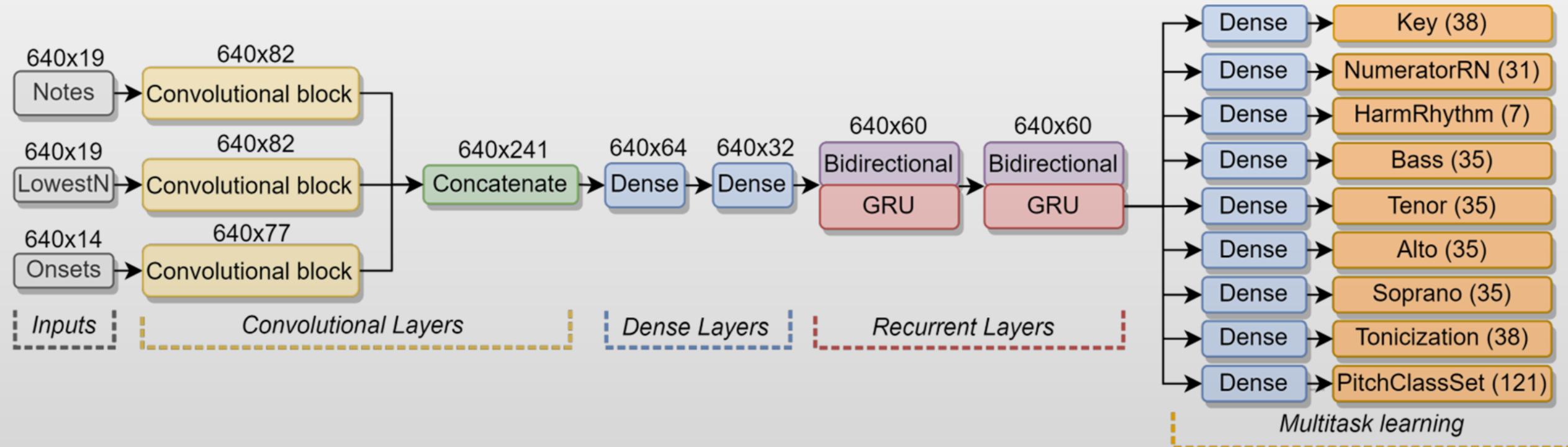




Contributions of my research

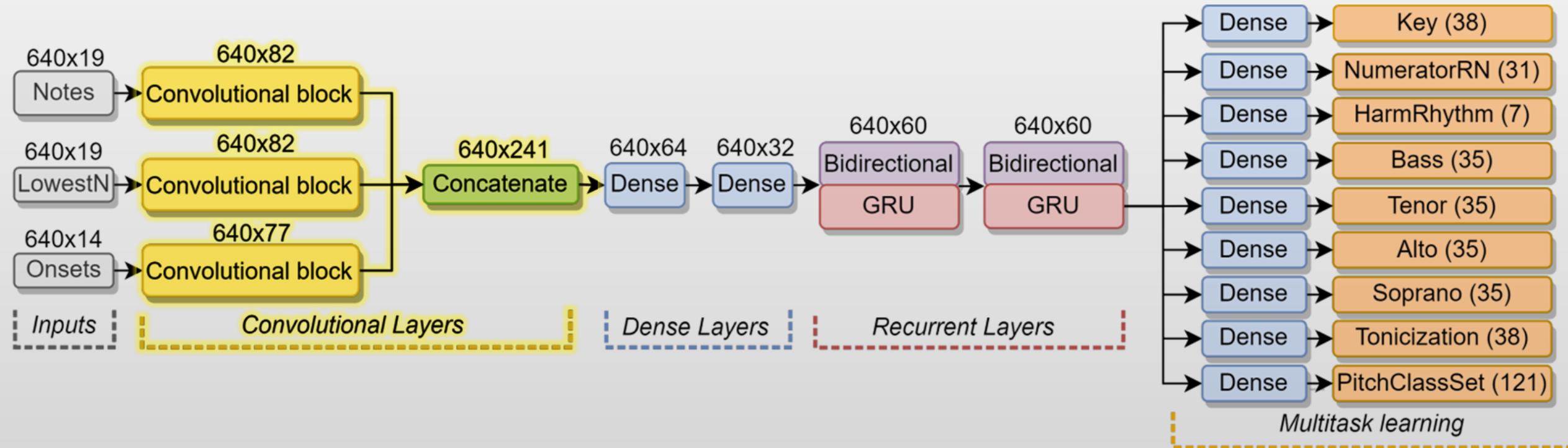
- A new neural network architecture, AugmentedNet

Neural Network Architecture



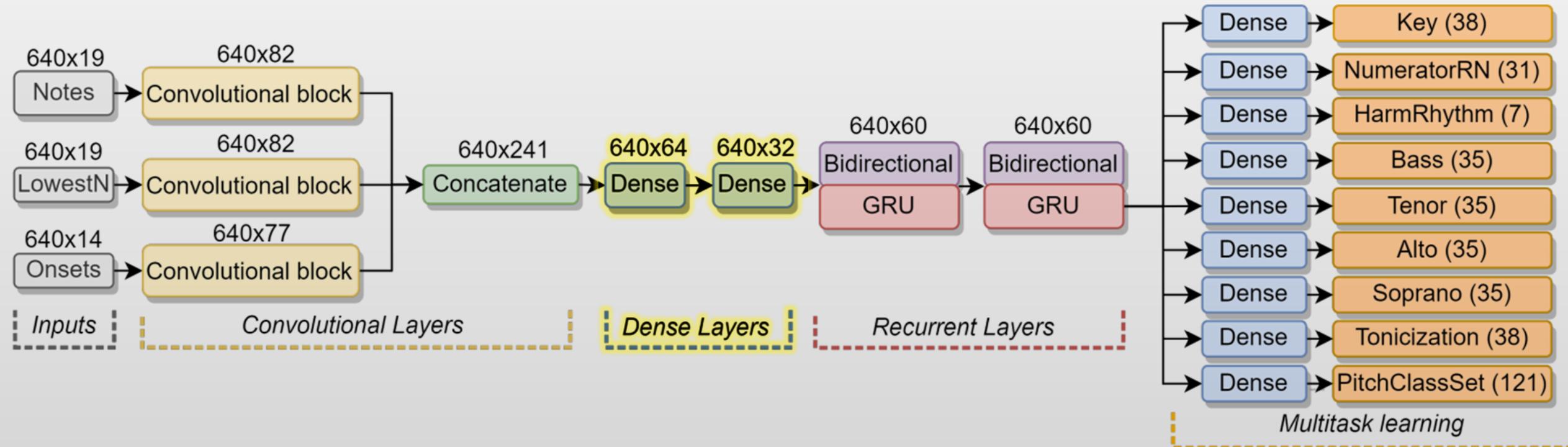
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



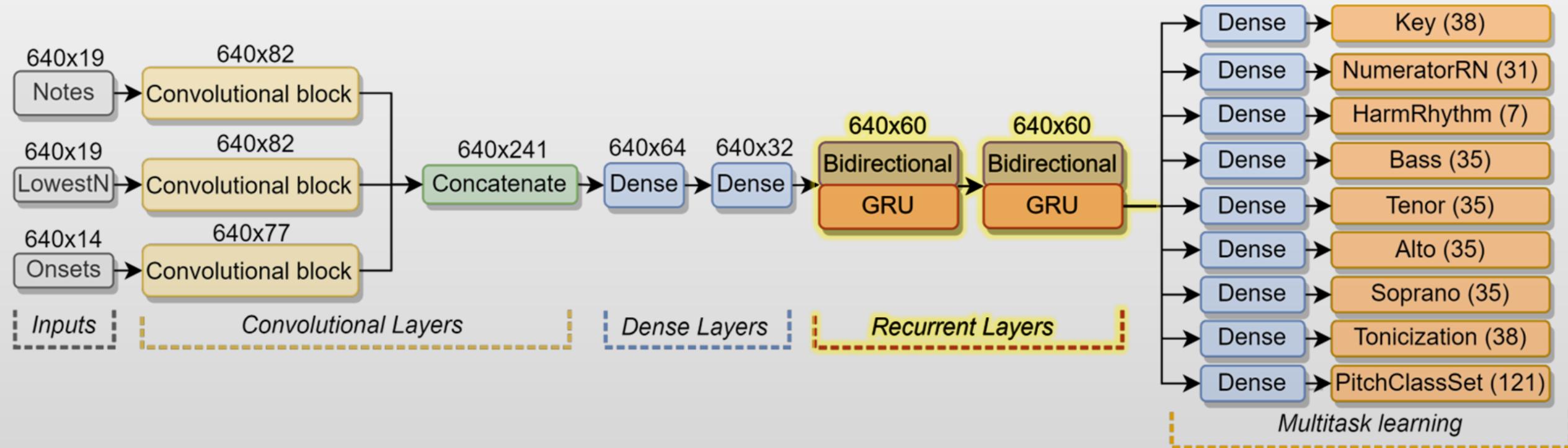
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



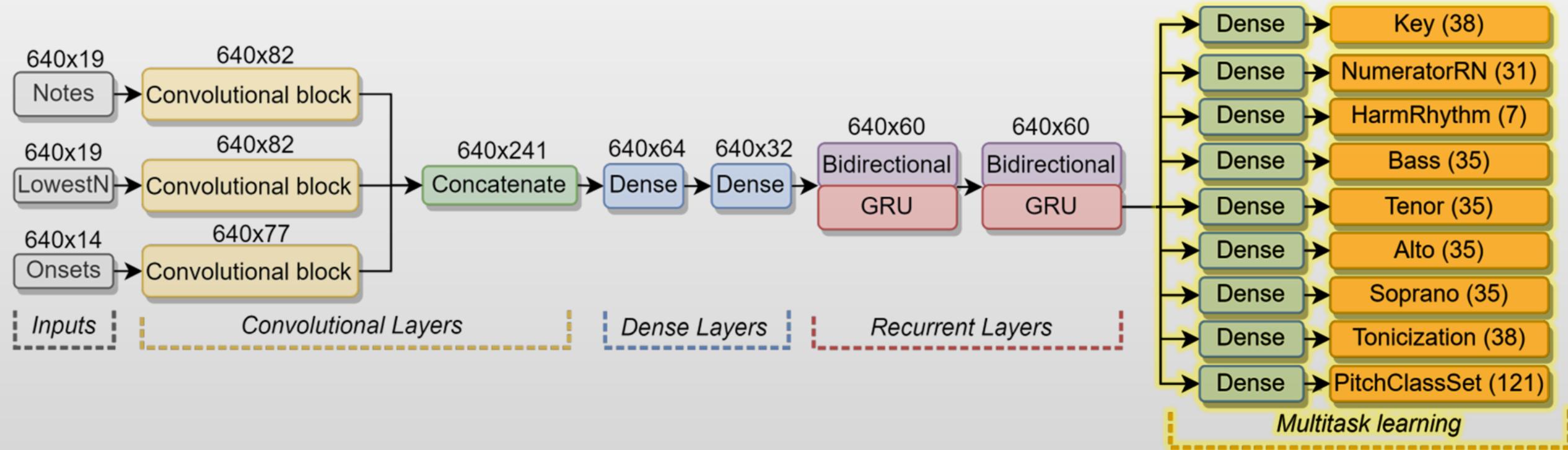
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



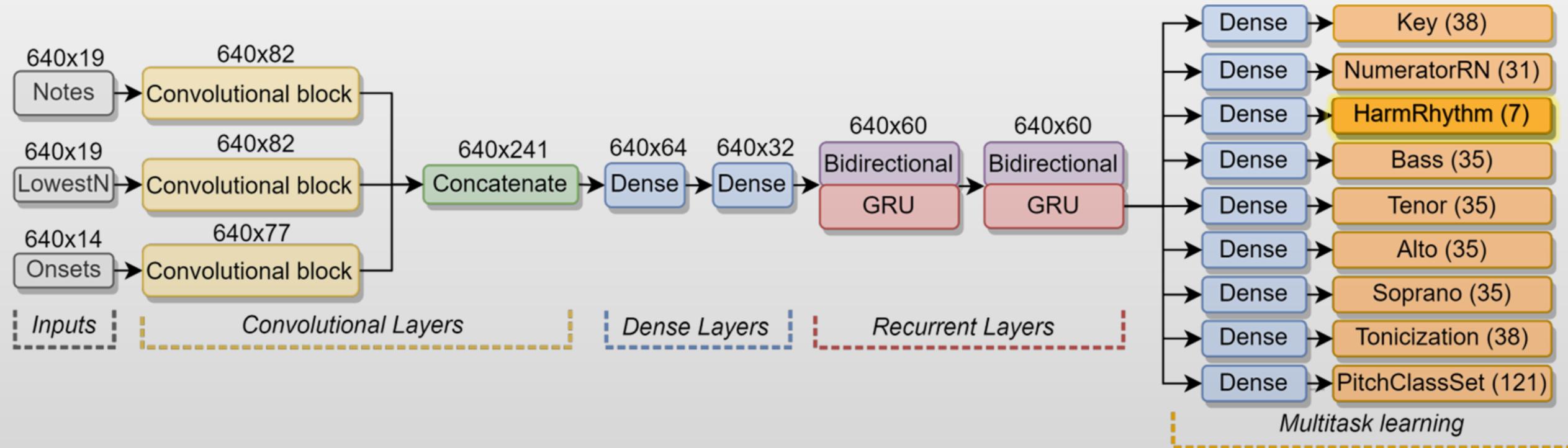
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



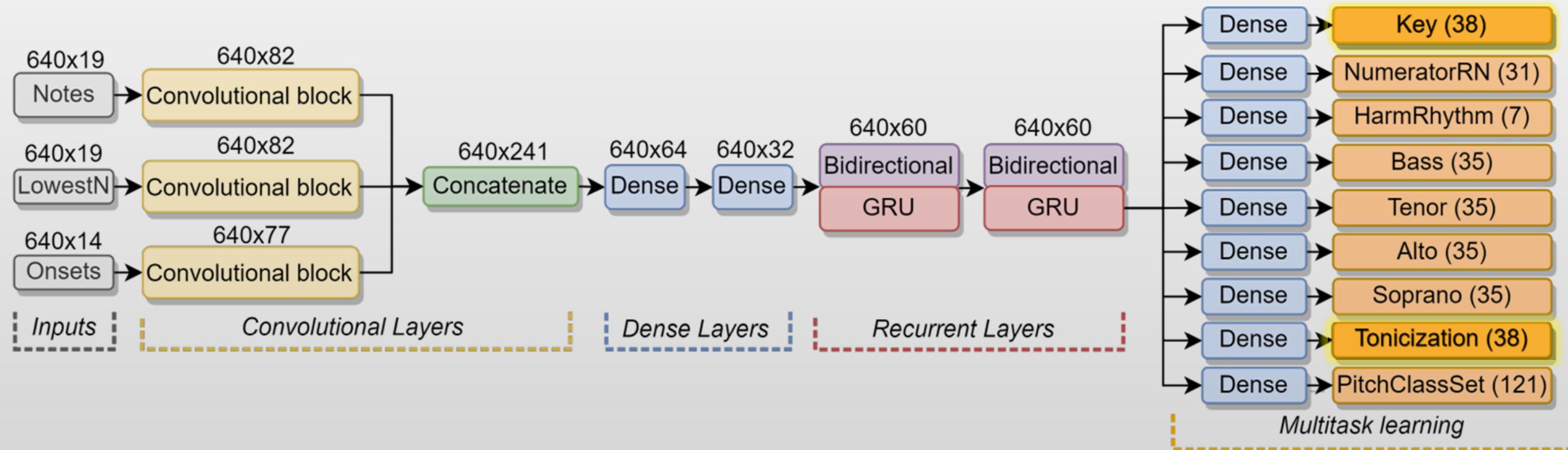
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



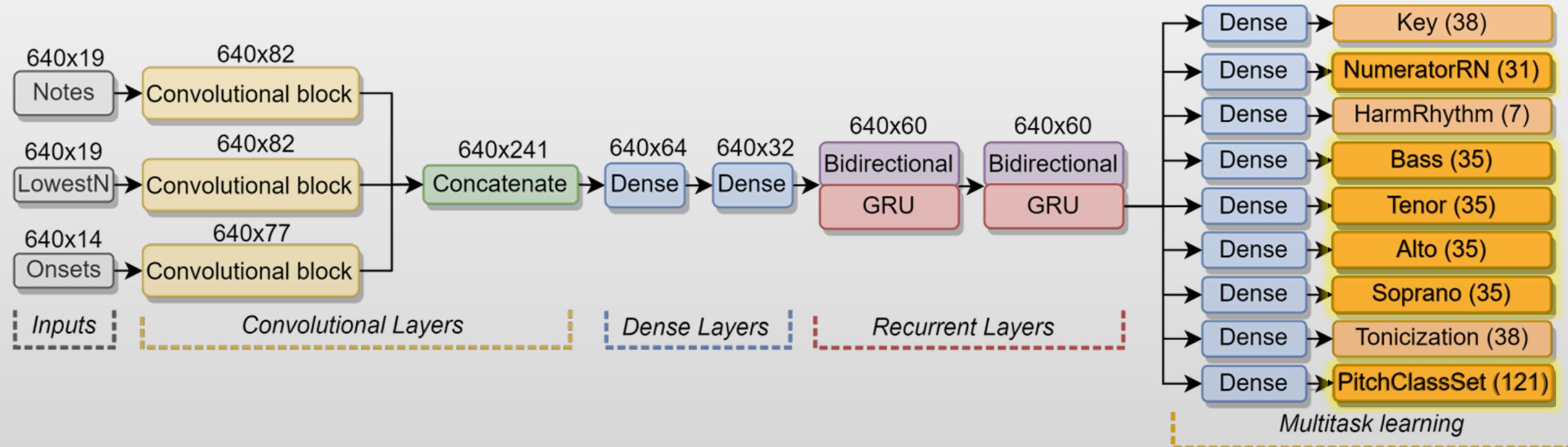
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



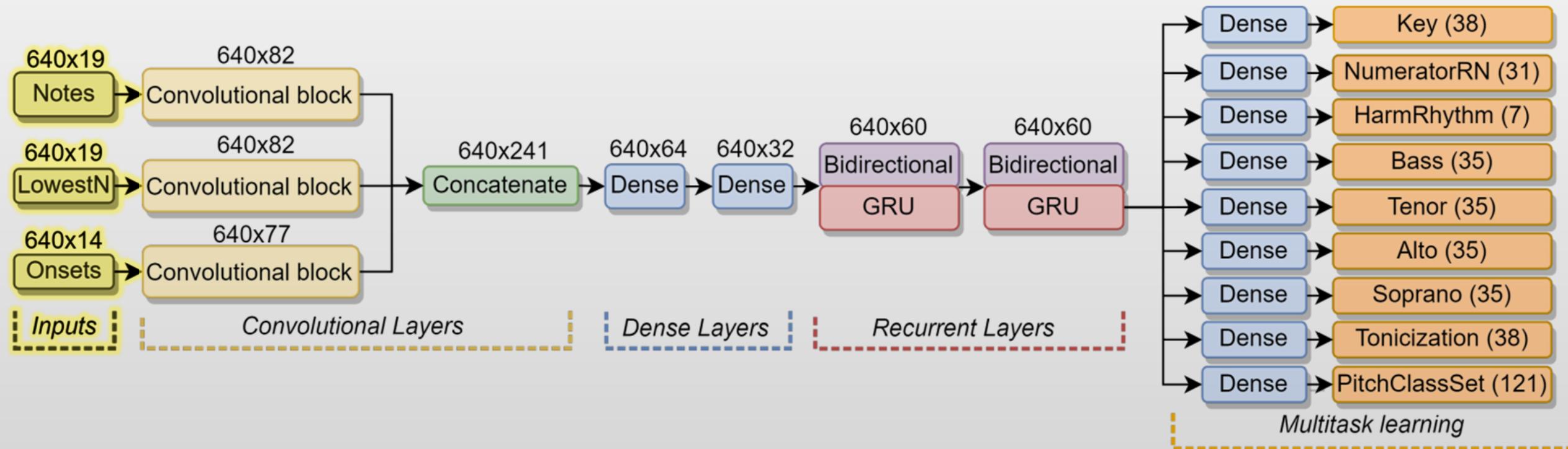
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture

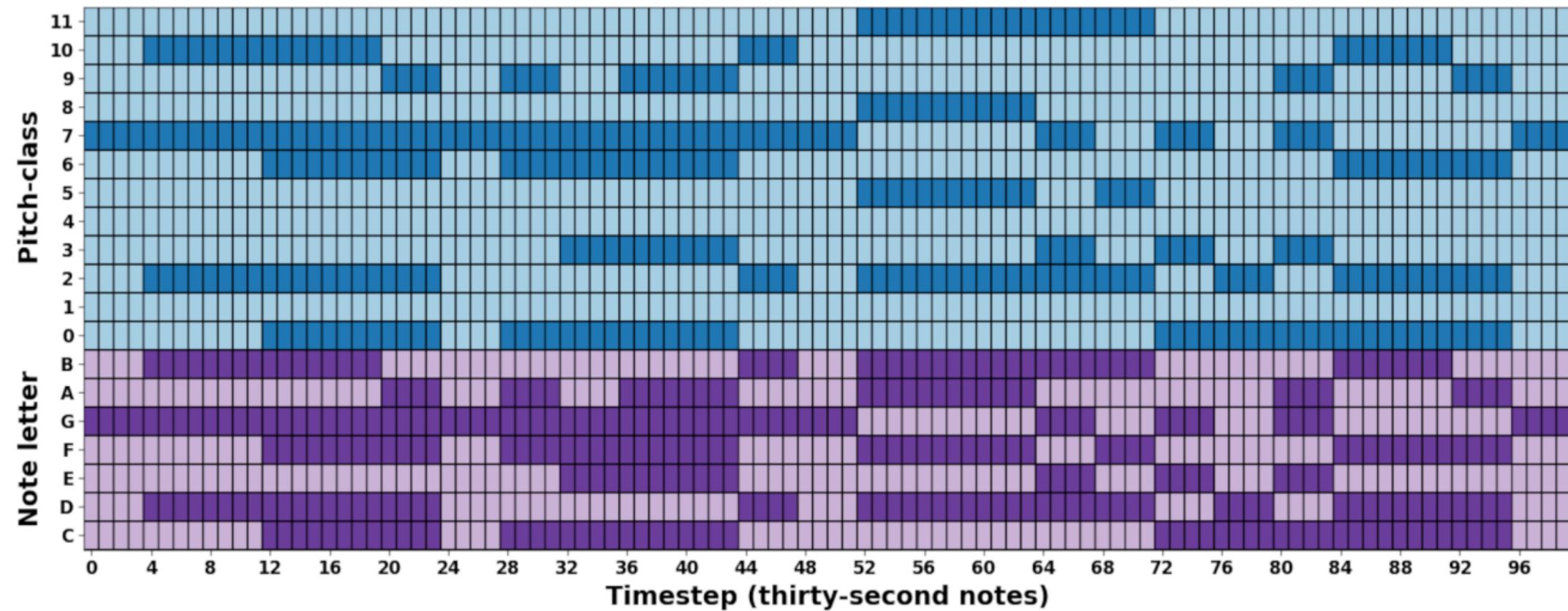


Convolutional Recurrent Neural Network (**CRNN**)

Input Components

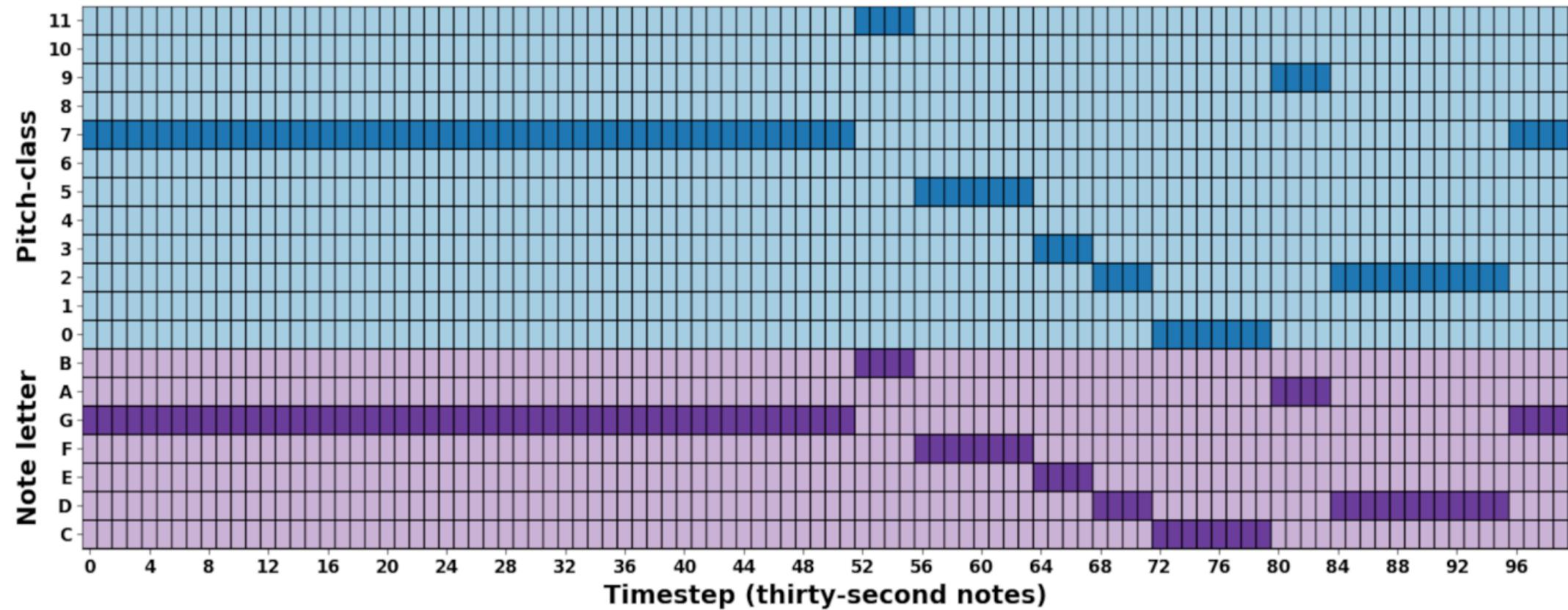
- All the sounding notes (**Notes19**)
 - All the pitches (with spelling) that are “sounding” at the current timestep
 - No octave information

A musical score for piano, featuring two staves. The top staff is in G minor (two flats) and 6/8 time, starting with a dynamic 'p'. The bottom staff is in C major (no sharps or flats) and 8/8 time. The music consists of eighth-note patterns and sustained notes.



Input Components

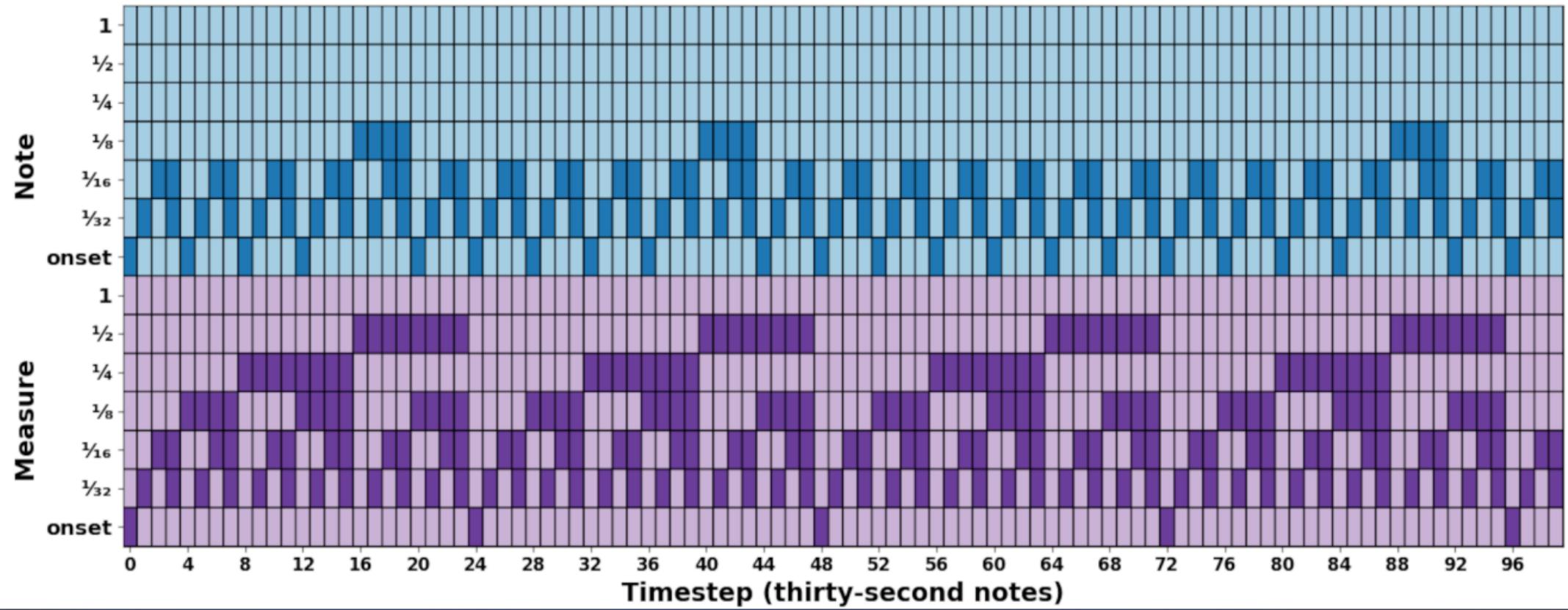
- All the sounding notes (**Notes19**)
 - All the pitches (with spelling) that are “sounding” at the current timestep
 - No octave information
- Lowest-sounding note (**LowestNote19**)
 - Similar to **Notes19**, but only the lowest-sounding note at the current timestep



Input Components

- All the sounding notes (**Notes19**)
 - All the pitches (with spelling) that are “sounding” at the current timestep
 - No octave information
- Lowest-sounding note (**LowestNote19**)
 - Similar to **Notes19**, but only the lowest-sounding note at the current timestep
- Onsets (**Onsets14**)
 - The location of note onsets (attacks) as well as the location of changes of measure

A musical score for piano in 6/8 time, featuring two staves. The top staff uses a treble clef and the bottom staff uses a bass clef. The key signature changes between measures, starting with one flat in the first measure and adding sharps in subsequent measures. Measure 11 begins with a dynamic *p*. Measures 12 and 13 show eighth-note patterns with grace notes. Measures 14 and 15 continue the rhythmic pattern, with measure 15 concluding with another dynamic *p*.



Ablation Studies

- Ablation studies are useful to inspect the contribution of the different components of a neural network architecture
- These studies refer to experiments where components of the model are removed or modified
- The purpose is to observe the effects that the modifications have in the performance of the model

Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

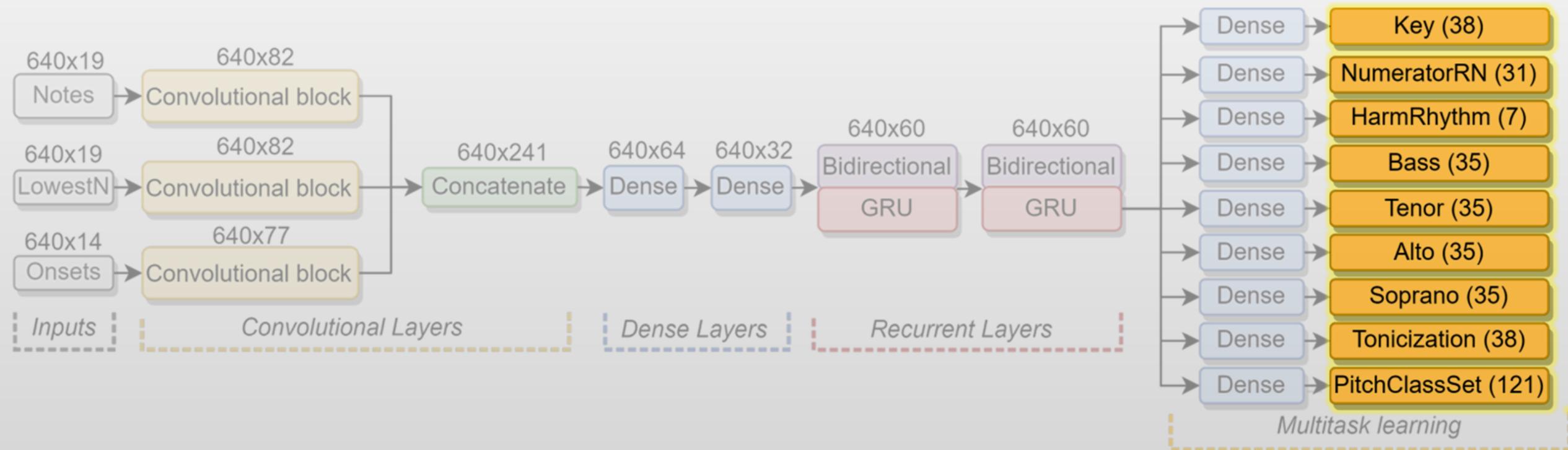
Average accuracy performance in a 5-fold cross-validation experiment

Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Average accuracy performance in a 5-fold cross-validation experiment

Neural Network Architecture



Convolutional Recurrent Neural Network (**CRNN**)

Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Average accuracy performance in a 5-fold cross-validation experiment

Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Average accuracy performance in a **5-fold cross-validation** experiment

Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Average accuracy performance in a **5-fold cross-validation** experiment

Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

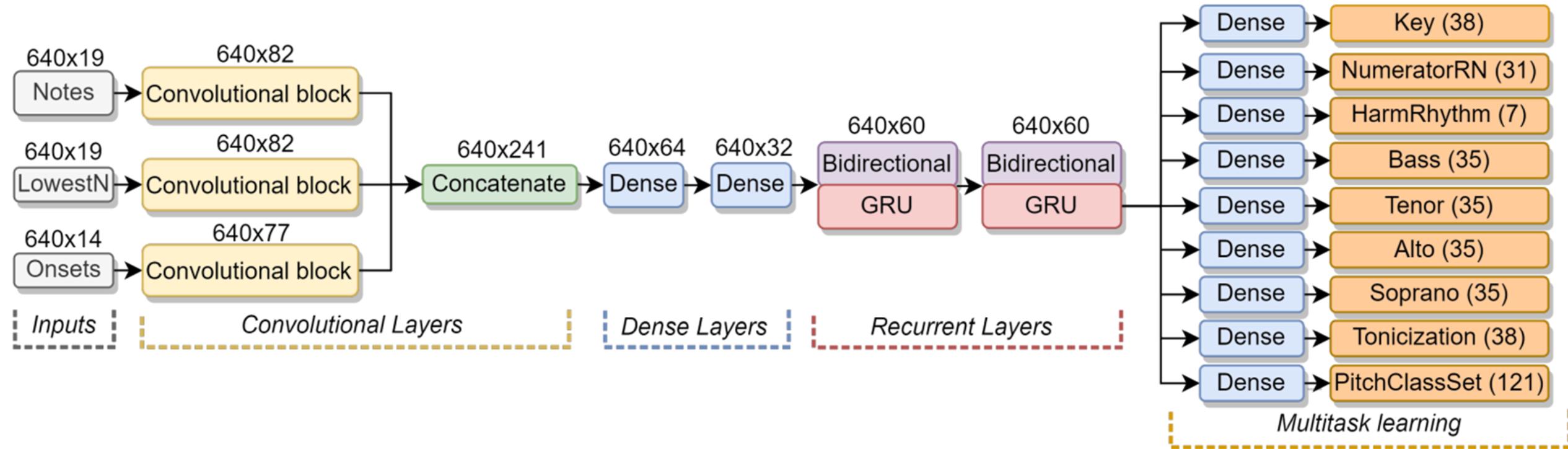
Average accuracy performance in a **5-fold cross-validation** experiment

Ablation Studies

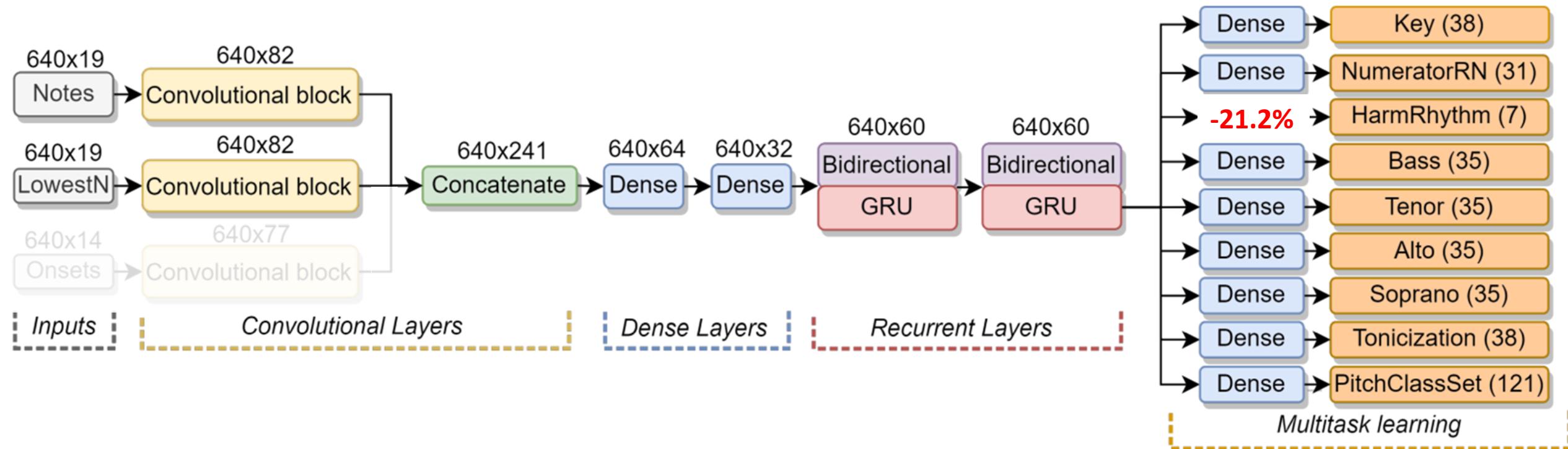
Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Average accuracy performance in a **5-fold cross-validation** experiment

Neural Network Architecture



Neural Network Architecture



Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

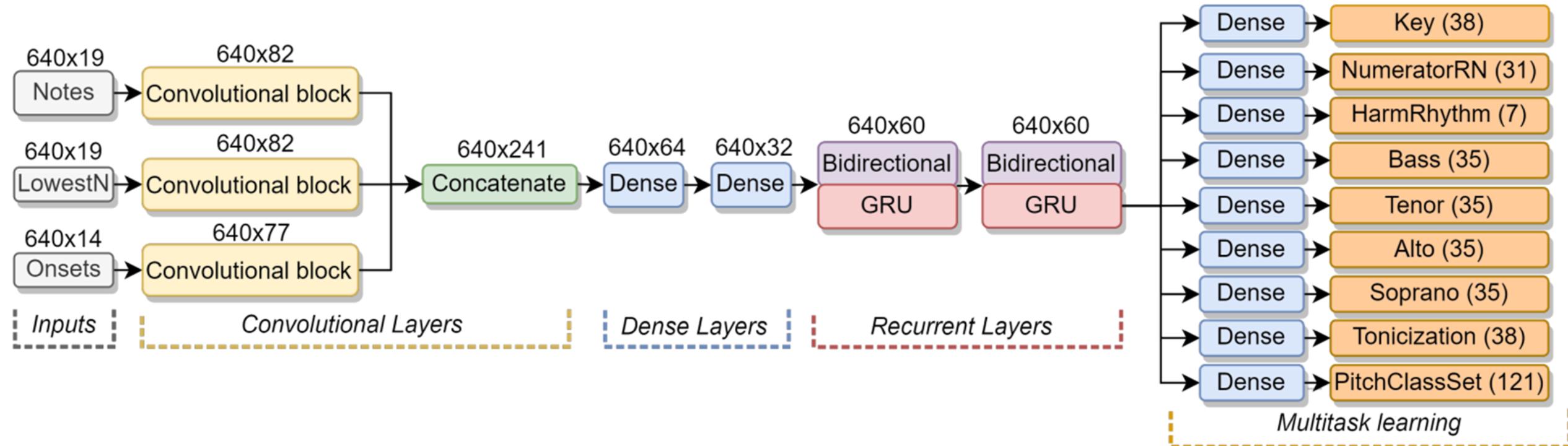
Average accuracy performance in a **5-fold cross-validation** experiment

Ablation Studies

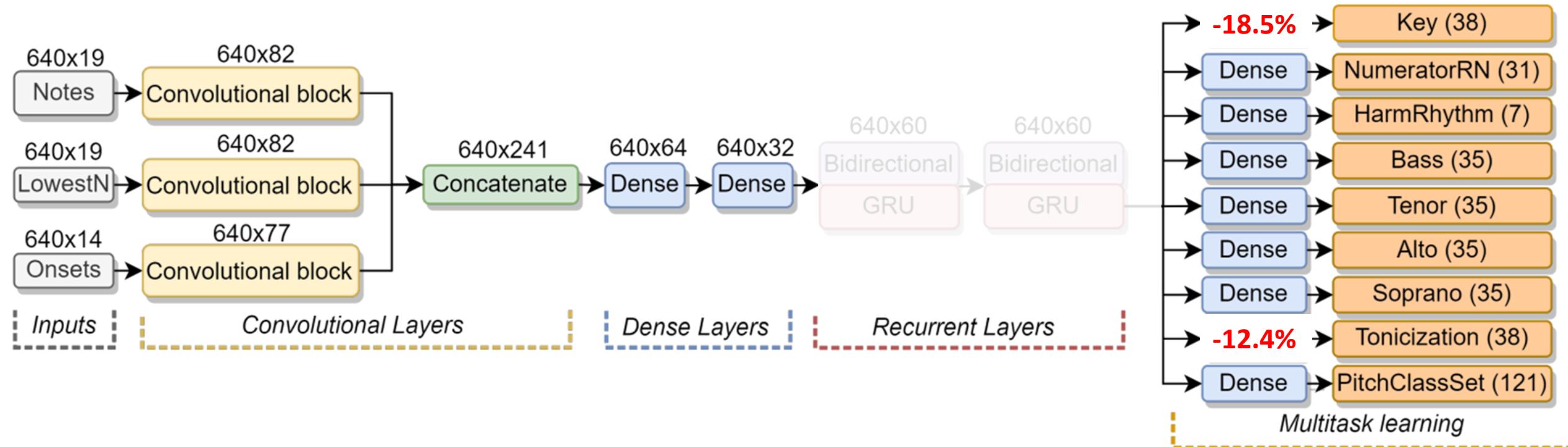
Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Average accuracy performance in a **5-fold cross-validation** experiment

Neural Network Architecture



Neural Network Architecture



Ablation Studies

Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	68.1 _{2.2}	74.5 _{2.2}	78.9 _{0.9}	78.8 _{3.4}	73.5 _{1.9}	61.3 _{2.8}	72.3 _{1.7}	71.3 _{2.5}	80 _{1.8}
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No LowestNote19	-13.1	-16.5	0.0	+0.3	+0.3	+0.4	-5.0	-15.4	+0.3
No Notes19	-11.1	-1.7	-4.7	-6.4	-22.2	-18.2	-16.7	-10.3	-10.0
No Onsets14	-1.4	-2.0	-21.2	+0.6	-1.1	-0.1	-0.8	-1.7	+0.7
Single block	-0.3	-0.4	+0.1	-0.4	+0.1	-0.2	0.0	-0.2	0.0
Constant filters	-0.3	-0.2	0.0	-0.5	-0.4	-1.0	-0.6	-0.2	-0.3
No convolutional	-0.4	-0.3	-0.3	+0.1	0.0	-0.4	-0.1	-0.1	+0.8
Linear dense	+0.2	+0.5	+0.4	-0.5	+0.8	+0.3	+0.7	+0.3	+0.1
No recurrent	-5.5	-5.0	-7.0	-18.5	-6.0	-10.9	-5.0	-6.2	-12.4
Unidirectional	-1.9	-1.1	-0.6	-6.7	-2.7	-5.2	-2.4	-2.0	-5.2

Average accuracy performance in a **5-fold cross-validation** experiment

Contributions of my research

- A new neural network architecture
- An aggregated dataset of Roman numeral annotations

Dataset(s)

Dataset

1. Annotated Beethoven Corpus ([ABC](#))
2. Beethoven Piano Sonatas ([BPS](#))
3. Haydn “Sun” String Quartets ([HaydnSun](#))
4. Key Modulations and Tonicizations Dataset ([KMT](#))
5. Mozart Piano Sonatas ([MPS](#))
6. Theme and Variations Encodings with Roman numerals ([TAVERN](#))
7. When in Rome ([WiR](#))

Published on

- Neuwirth et al. (2018)
- Chen and Su (2018)
- Nápoles López (2017)
- Nápoles López et al. (2020)
- Hentschel et al. (2021)
- Devaney et al. (2015)
- Gotham, Tymoczko, and Cuthbert (2019)

Dataset(s)

Aggregated dataset (7 publicly available datasets)

- 58,393 measures of music
- 175,930 quarter notes in duration
- 104,926 Roman numeral analysis annotations

Divided into **training**, **validation**, and **test** sets to conduct experiments

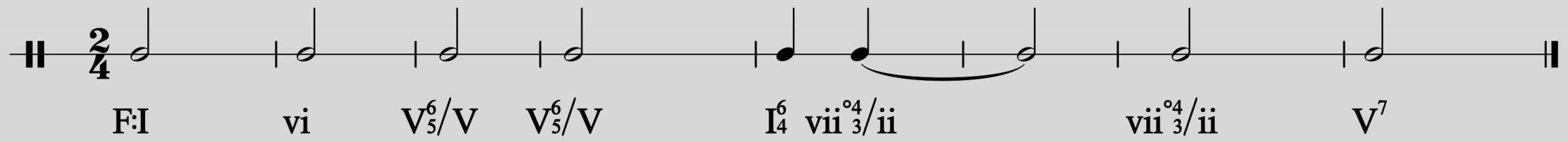
Data Augmentation - Transposition

- 100,000 chords is not a very large dataset for deep learning
- The most common way of doing data augmentation in automatic Roman numeral analysis models is via **transposition**

Contributions of my research

- A new neural network architecture
- An aggregated dataset of Roman numeral annotations
- A new data-augmentation technique

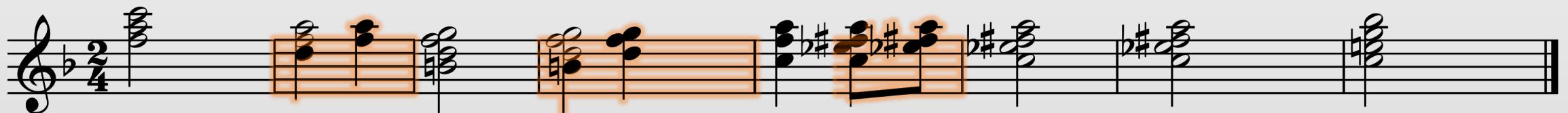
Data Augmentation by Synthesis



Data Augmentation by Synthesis

A musical score and harmonic analysis diagram. The top part shows a staff with a treble clef, a key signature of one flat, and a time signature of 2/4. It consists of eight measures of chords: F major (two measures), C major (one measure), G major (one measure), G major (one measure), A major (one measure), D major (one measure), D major (one measure), and E major (one measure). The bottom part shows a horizontal timeline with vertical bar lines corresponding to the measures above. Below the timeline, harmonic labels are placed under each measure: F:I, vi, V⁶/₅/V, V⁶/₅/V, I⁶₄, vii^{°4}₃/ii, vii^{°4}₃/ii, and V⁷. Measure 6 includes a bracket under the first two notes of the measure, indicating a harmonic progression from I⁶₄ to vii^{°4}₃/ii.

Data Augmentation by Synthesis



Separate the bass

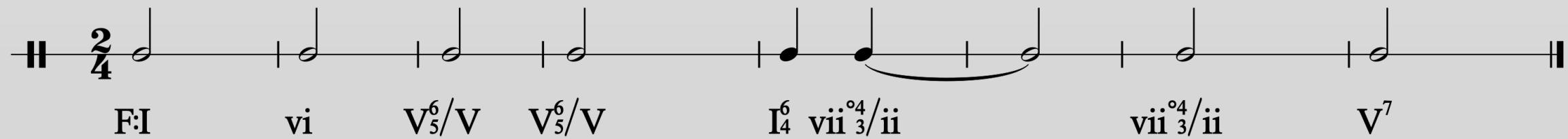
2/4

F:I vi V⁶/V V⁶/V I⁶₄ vii^{°4}₃/ii vii^{°4}₃/ii V⁷

Data Augmentation by Synthesis



Syncopation



Data Augmentation by Synthesis

A musical score consisting of two parts. The top part shows a treble clef staff in 2/4 time with a blue Alberti bass line highlighted. The bottom part shows a harmonic analysis timeline with measures labeled: F:I, vi, V⁶/₅/V, V⁶/₅/V, I⁶₄, vii^{°4}₃/ii, vii^{°4}₃/ii, and V⁷.

Alberti bass

2/4

F:I vi V⁶/₅/V V⁶/₅/V I⁶₄ vii^{°4}₃/ii vii^{°4}₃/ii V⁷

Data Augmentation by Synthesis

A musical score and its corresponding harmonic analysis diagram. The score consists of two staves. The top staff is a treble clef staff with a key signature of one flat, indicating F major. It shows a sequence of notes and rests. The bottom staff is a bass clef staff with a key signature of one flat, also indicating F major. It shows a sequence of notes and rests. Below the staffs is a harmonic analysis diagram consisting of a horizontal line with vertical tick marks. Above the first tick mark is the number '2' above a '4'. Below the staff, under the first tick mark, is the label 'F:I'. Between the first and second tick marks, there is a vertical bar line. Below the second tick mark is the label 'vi'. Between the second and third tick marks, there is a vertical bar line. Below the third tick mark is the label 'V₅⁶/V'. Between the third and fourth tick marks, there is a vertical bar line. Below the fourth tick mark is the label 'V₅⁶/V'. Between the fourth and fifth tick marks, there is a vertical bar line. Below the fifth tick mark is the label 'I₄⁶ vii^{°4}₃/ii'. Between the fifth and sixth tick marks, there is a vertical bar line. Below the sixth tick mark is a note connected by a curved line to a note below the seventh tick mark, which is labeled 'vii^{°4}₃/ii'. Between the seventh and eighth tick marks, there is a vertical bar line. Below the eighth tick mark is the label 'V⁷'. After the eighth tick mark, the horizontal line continues without a staff.

Effects of Data Augmentation

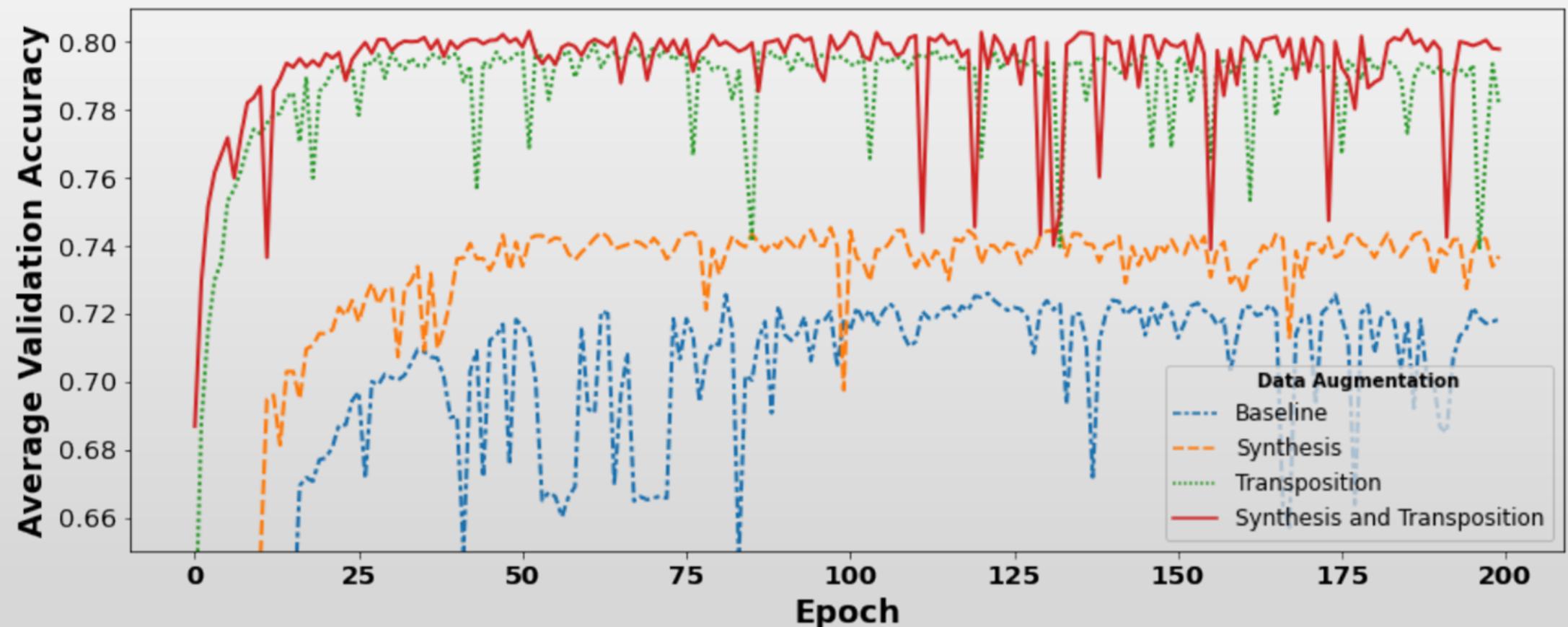


Figure 6.14: Average validation accuracy achieved in the four experiments with data augmentation. The average accuracy value is across the seven datasets at the given epoch.

Effects of Data Augmentation

Table 6.7: *Average performance and standard deviation of the four experiments with data augmentation over the seven publicly available datasets.*

	No augmentation	Synthesis	Transposition	Synthesis and Transposition
ABC	65.2 _{7.04}	67.2 _{7.00}	70.7 _{6.09}	72.2 _{5.71}
BPS	62.5 _{9.25}	64.6 _{6.62}	74.7 _{5.89}	75.7 _{5.70}
HaydnSun	54.4 _{11.76}	59.4 _{9.64}	68.2 _{8.61}	68.7 _{6.87}
KMT	78.8 _{6.02}	81.6 _{6.65}	85.3 _{5.84}	88.7 _{3.60}
MPS	73.2 _{7.52}	75.2 _{6.86}	81.8 _{5.90}	81.5 _{5.89}
TAVERN	69.1 _{7.55}	71.3 _{7.66}	76.8 _{6.07}	77.0 _{6.44}
WiR	69.9 _{8.79}	71.7 _{7.88}	78.1 _{5.40}	78.6 _{4.74}

Effects of Data Augmentation

Table 6.7: *Average performance and standard deviation of the four experiments with data augmentation over the seven publicly available datasets.*

	No augmentation	Synthesis	Transposition	Synthesis and Transposition
ABC	65.2 _{7.04}	67.2 _{7.00}	70.7 _{6.09}	72.2 _{5.71}
BPS	62.5 _{9.25}	64.6 _{6.62}	74.7 _{5.89}	75.7 _{5.70}
HaydnSun	54.4 _{11.76}	59.4 _{9.64}	68.2 _{8.61}	68.7 _{6.87}
KMT	78.8 _{6.02}	81.6 _{6.65}	85.3 _{5.84}	88.7 _{3.60}
MPS	73.2 _{7.52}	75.2 _{6.86}	81.8 _{5.90}	81.5 _{5.89}
TAVERN	69.1 _{7.55}	71.3 _{7.66}	76.8 _{6.07}	77.0 _{6.44}
WiR	69.9 _{8.79}	71.7 _{7.88}	78.1 _{5.40}	78.6 _{4.74}

Contributions of my research

- A new neural network architecture
- An aggregated dataset of Roman numeral annotations
- A new data-augmentation technique
- An evaluation of the proposed model compared to previous ones

Contributions of my research

- A new neural network architecture
- An aggregated dataset of Roman numeral annotations
- A new data-augmentation technique
- An evaluation of the proposed model compared to previous ones
 - Time performance
 - Accuracy on individual tasks
 - Accuracy on individual Roman numeral classes (sorted by occurrence)

Models Evaluated

- Melisma (2003)
- Chen and Su (2021)
- Micchi et al. (2021)
- McLeod and Rohrmeier (2021)
- Nápoles López (2022)
 - Referred to as “AugmentedNet”

Results obtained

Time Performance

Table 6.11: *Time elapsed for each model to provide the output predictions on the 94 files of the test set.*

Model	Total time	Mean	SD
Melisma	2.7 min	1.7s	0.8s
Micchi et al. (2021)	18.6 min	11.9s	6.0s
AugmentedNet	22.6 min	14.4s	12.5s
Chen and Su (2021)	26.2 min	16.8s	9.0s
McLeod and Rohrmeier (2021)	183.0 min	116.8s	154.6s

Accuracy on Individual Tasks

Table 6.12: *Comparison of the accuracy achieved by the five models on the individual tasks ρ , κ , and ι .*

model	pcset (ρ)	Key (κ)	Inversion (ι)
Melisma	54.6	58.2	70.3
McLeod and Rohrmeier (2021)	57.7	51.5	69.3
Chen and Su (2021)	59.8	52.6	70.8
Micchi et al. (2021)	74.7	72.9	82.0
AugmentedNet	79.4	79.4	81.9

(pcset refers to a chord label expressed as a pitch-class set)

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
III⁺⁷	0.02	0	0	0	0	0
Fr⁷	0.03	0	0	0	0	57.9
III⁺	0.04	0	0	4.1	8.3	41.7
i⁷	0.06	19.0	0	0	23.2	0
I⁷	0.07	0	0	0	0	0
V⁺	0.09	0	12.8	10.6	23.4	40.4
IV⁷	0.09	0	0	0	0	0
iii⁷	0.10	5.4	22.2	0	0	3.7
iv⁷	0.16	33.3	7	20.0	22.2	37.8
vi⁷	0.19	3.2	29.3	1.0	0	15.6
VI⁷	0.19	0	0	0	7.7	15.3
It	0.22	0	24.3	0	0	32.4
vii^{o7}	0.25	0	7.3	0	1.5	14.3
Ger⁷	0.43	0	0	25.7	0.9	44.6
ii^o	0.47	0.6	14.6	26.0	13.5	35.0
N	0.51	0	43.4	64.2	35.0	32.0
ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
iii	0.78	18.7	2.0	27.0	27.8	23.2
ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
Cad₄⁶	1.13	0	0	0	0	6.9
VI	1.40	5.5	1.6	25.0	36.0	47.4
vi	1.82	21.6	28.5	31.2	28.2	35.2
iv	1.88	28.1	29	56.6	37.2	51.4
vii^o	2.37	5.3	21.9	11.4	25.2	43.8
ii	3.70	27.4	45.5	54.9	33.0	60.3
IV	3.78	28.0	50.5	58.9	26.7	57.0
vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
i	10.93	49.4	71.6	78.8	56.0	82.3
V	13.89	37.6	44.9	51.9	56.3	60.9
V⁷	23.17	43.1	57.9	68.0	55.9	68.6
I	25.18	60.5	75.7	83.1	59.5	84.3

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
III⁺⁷	0.02	0	0	0	0	0
Fr⁷	0.03	0	0	0	0	57.9
III⁺	0.04	0	0	4.1	8.3	41.7
i⁷	0.06	19.0	0	0	23.2	0
I⁷	0.07	0	0	0	0	0
V⁺	0.09	0	12.8	10.6	23.4	40.4
IV⁷	0.09	0	0	0	0	0
iii⁷	0.10	5.4	22.2	0	0	3.7
iv⁷	0.16	33.3	7	20.0	22.2	37.8
vi⁷	0.19	3.2	29.3	1.0	0	15.6
VI⁷	0.19	0	0	0	7.7	15.3
It	0.22	0	24.3	0	0	32.4
vii^{o7}	0.25	0	7.3	0	1.5	14.3
Ger⁷	0.43	0	0	25.7	0.9	44.6
ii^o	0.47	0.6	14.6	26.0	13.5	35.0
N	0.51	0	43.4	64.2	35.0	32.0
ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
iii	0.78	18.7	2.0	27.0	27.8	23.2
ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
Cad₄⁶	1.13	0	0	0	0	6.9
VI	1.40	5.5	1.6	25.0	36.0	47.4
vi	1.82	21.6	28.5	31.2	28.2	35.2
iv	1.88	28.1	29	56.6	37.2	51.4
vii^o	2.37	5.3	21.9	11.4	25.2	43.8
ii	3.70	27.4	45.5	54.9	33.0	60.3
IV	3.78	28.0	50.5	58.9	26.7	57.0
vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
i	10.93	49.4	71.6	78.8	56.0	82.3
v	13.89	37.6	44.9	51.9	56.3	60.9
v⁷	23.17	43.1	57.9	68.0	55.9	68.6
I	25.18	60.5	75.7	83.1	59.5	84.3

73% of the Roman
numeral labels

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
III⁺⁷	0.02	0	0	0	0	0
Fr⁷	0.03	0	0	0	0	57.9
III⁺	0.04	0	0	4.1	8.3	41.7
i⁷	0.06	19.0	0	0	23.2	0
I⁷	0.07	0	0	0	0	0
V⁺	0.09	0	12.8	10.6	23.4	40.4
IV⁷	0.09	0	0	0	0	0
iii⁷	0.10	5.4	22.2	0	0	3.7
iv⁷	0.16	33.3	7	20.0	22.2	37.8
vi⁷	0.19	3.2	29.3	1.0	0	15.6
VI⁷	0.19	0	0	0	7.7	15.3
It	0.22	0	24.3	0	0	32.4
vii^{o7}	0.25	0	7.3	0	1.5	14.3
Ger⁷	0.43	0	0	25.7	0.9	44.6
ii^o	0.47	0.6	14.6	26.0	13.5	35.0
N	0.51	0	43.4	64.2	35.0	32.0
ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
iii	0.78	18.7	2.0	27.0	27.8	23.2
ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
Cad₄⁶	1.13	0	0	0	0	6.9
VI	1.40	5.5	1.6	25.0	36.0	47.4
vi	1.82	21.6	28.5	31.2	28.2	35.2
iv	1.88	28.1	29	56.6	37.2	51.4
vii^o	2.37	5.3	21.9	11.4	25.2	43.8
ii	3.70	27.4	45.5	54.9	33.0	60.3
IV	3.78	28.0	50.5	58.9	26.7	57.0
vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
i	10.93	49.4	71.6	78.8	56.0	82.3
V	13.89	37.6	44.9	51.9	56.3	60.9
V⁷	23.17	43.1	57.9	68.0	55.9	68.6
I	25.18	60.5	75.7	83.1	59.5	84.3

Melisma (2003)

0 out of 31 classes

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
III⁺⁷	0.02	0	0	0	0	0
Fr⁷	0.03	0	0	0	0	57.9
III⁺	0.04	0	0	4.1	8.3	41.7
i⁷	0.06	19.0	0	0	23.2	0
I⁷	0.07	0	0	0	0	0
V⁺	0.09	0	12.8	10.6	23.4	40.4
IV⁷	0.09	0	0	0	0	0
iii⁷	0.10	5.4	22.2	0	0	3.7
iv⁷	0.16	33.3	7	20.0	22.2	37.8
vi⁷	0.19	3.2	29.3	1.0	0	15.6
VI⁷	0.19	0	0	0	7.7	15.3
It	0.22	0	24.3	0	0	32.4
vii^{o7}	0.25	0	7.3	0	1.5	14.3
Ger⁷	0.43	0	0	25.7	0.9	44.6
ii^o	0.47	0.6	14.6	26.0	13.5	35.0
N	0.51	0	43.4	64.2	35.0	32.0
ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
iii	0.78	18.7	2.0	27.0	27.8	23.2
ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
Cad₄⁶	1.13	0	0	0	0	6.9
VI	1.40	5.5	1.6	25.0	36.0	47.4
vi	1.82	21.6	28.5	31.2	28.2	35.2
iv	1.88	28.1	29	56.6	37.2	51.4
vii^o	2.37	5.3	21.9	11.4	25.2	43.8
ii	3.70	27.4	45.5	54.9	33.0	60.3
IV	3.78	28.0	50.5	58.9	26.7	57.0
vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
i	10.93	49.4	71.6	78.8	56.0	82.3
V	13.89	37.6	44.9	51.9	56.3	60.9
V⁷	23.17	43.1	57.9	68.0	55.9	68.6
I	25.18	60.5	75.7	83.1	59.5	84.3

Melisma (2003)

0 out of 31 classes

Chen and Su (2021)

2 out of 31 classes

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
III⁺⁷	0.02	0	0	0	0	0
Fr⁷	0.03	0	0	0	0	57.9
III⁺	0.04	0	0	4.1	8.3	41.7
i⁷	0.06	19.0	0	0	23.2	0
I⁷	0.07	0	0	0	0	0
V⁺	0.09	0	12.8	10.6	23.4	40.4
IV⁷	0.09	0	0	0	0	0
iii⁷	0.10	5.4	22.2	0	0	3.7
iv⁷	0.16	33.3	7	20.0	22.2	37.8
vi⁷	0.19	3.2	29.3	1.0	0	15.6
VI⁷	0.19	0	0	0	7.7	15.3
It	0.22	0	24.3	0	0	32.4
vii^{o7}	0.25	0	7.3	0	1.5	14.3
Ger⁷	0.43	0	0	25.7	0.9	44.6
ii^o	0.47	0.6	14.6	26.0	13.5	35.0
N	0.51	0	43.4	64.2	35.0	32.0
ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
iii	0.78	18.7	2.0	27.0	27.8	23.2
ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
Cad₄⁶	1.13	0	0	0	0	6.9
VI	1.40	5.5	1.6	25.0	36.0	47.4
vi	1.82	21.6	28.5	31.2	28.2	35.2
iv	1.88	28.1	29	56.6	37.2	51.4
vii^o	2.37	5.3	21.9	11.4	25.2	43.8
ii	3.70	27.4	45.5	54.9	33.0	60.3
IV	3.78	28.0	50.5	58.9	26.7	57.0
vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
i	10.93	49.4	71.6	78.8	56.0	82.3
V	13.89	37.6	44.9	51.9	56.3	60.9
V⁷	23.17	43.1	57.9	68.0	55.9	68.6
I	25.18	60.5	75.7	83.1	59.5	84.3

Melisma (2003)

0 out of 31 classes

Chen and Su (2021)

2 out of 31 classes

Micchi et al. (2021)

3 out of 31 classes

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
III⁺⁷	0.02	0	0	0	0	0
Fr⁷	0.03	0	0	0	0	57.9
III⁺	0.04	0	0	4.1	8.3	41.7
i⁷	0.06	19.0	0	0	23.2	0
I⁷	0.07	0	0	0	0	0
V⁺	0.09	0	12.8	10.6	23.4	40.4
IV⁷	0.09	0	0	0	0	0
iii⁷	0.10	5.4	22.2	0	0	3.7
iv⁷	0.16	33.3	7	20.0	22.2	37.8
vi⁷	0.19	3.2	29.3	1.0	0	15.6
VI⁷	0.19	0	0	0	7.7	15.3
It	0.22	0	24.3	0	0	32.4
vii^{o7}	0.25	0	7.3	0	1.5	14.3
Ger⁷	0.43	0	0	25.7	0.9	44.6
ii^o	0.47	0.6	14.6	26.0	13.5	35.0
N	0.51	0	43.4	64.2	35.0	32.0
ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
iii	0.78	18.7	2.0	27.0	27.8	23.2
ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
Cad₄⁶	1.13	0	0	0	0	6.9
VI	1.40	5.5	1.6	25.0	36.0	47.4
vi	1.82	21.6	28.5	31.2	28.2	35.2
iv	1.88	28.1	29	56.6	37.2	51.4
vii^o	2.37	5.3	21.9	11.4	25.2	43.8
ii	3.70	27.4	45.5	54.9	33.0	60.3
IV	3.78	28.0	50.5	58.9	26.7	57.0
vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
i	10.93	49.4	71.6	78.8	56.0	82.3
V	13.89	37.6	44.9	51.9	56.3	60.9
V⁷	23.17	43.1	57.9	68.0	55.9	68.6
I	25.18	60.5	75.7	83.1	59.5	84.3

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

	Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
Melisma (2003) 0 out of 31 classes	III⁺⁷	0.02	0	0	0	0	0
	Fr⁷	0.03	0	0	0	0	57.9
	III⁺	0.04	0	0	4.1	8.3	41.7
	i⁷	0.06	19.0	0	0	23.2	0
	I⁷	0.07	0	0	0	0	0
	V⁺	0.09	0	12.8	10.6	23.4	40.4
	IV⁷	0.09	0	0	0	0	0
	iii⁷	0.10	5.4	22.2	0	0	3.7
	iv⁷	0.16	33.3	7	20.0	22.2	37.8
	vi⁷	0.19	3.2	29.3	1.0	0	15.6
Chen and Su (2021) 2 out of 31 classes	VI⁷	0.19	0	0	0	7.7	15.3
	It	0.22	0	24.3	0	0	32.4
	vii^{o7}	0.25	0	7.3	0	1.5	14.3
	Ger⁷	0.43	0	0	25.7	0.9	44.6
	ii^o	0.47	0.6	14.6	26.0	13.5	35.0
	N	0.51	0	43.4	64.2	35.0	32.0
	ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
	iii	0.78	18.7	2.0	27.0	27.8	23.2
	ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
	Cad₄⁶	1.13	0	0	0	0	6.9
AugmentedNet 21 out of 31 classes	VI	1.40	5.5	1.6	25.0	36.0	47.4
	vi	1.82	21.6	28.5	31.2	28.2	35.2
	iv	1.88	28.1	29	56.6	37.2	51.4
	vii^o	2.37	5.3	21.9	11.4	25.2	43.8
	ii	3.70	27.4	45.5	54.9	33.0	60.3
	IV	3.78	28.0	50.5	58.9	26.7	57.0
	vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
	i	10.93	49.4	71.6	78.8	56.0	82.3
	V	13.89	37.6	44.9	51.9	56.3	60.9
	V⁷	23.17	43.1	57.9	68.0	55.9	68.6
	I	25.18	60.5	75.7	83.1	59.5	84.3

Table 6.13: Accuracy performance of the models in all 31 numerator classes, sorted by least occurrence in the test set.

Melisma (2003)

0 out of 31 classes

Chen and Su (2021)

2 out of 31 classes

Micchi et al. (2021)

3 out of 31 classes

McLeod and Rohrmeier (2021)

2 out of 31 classes

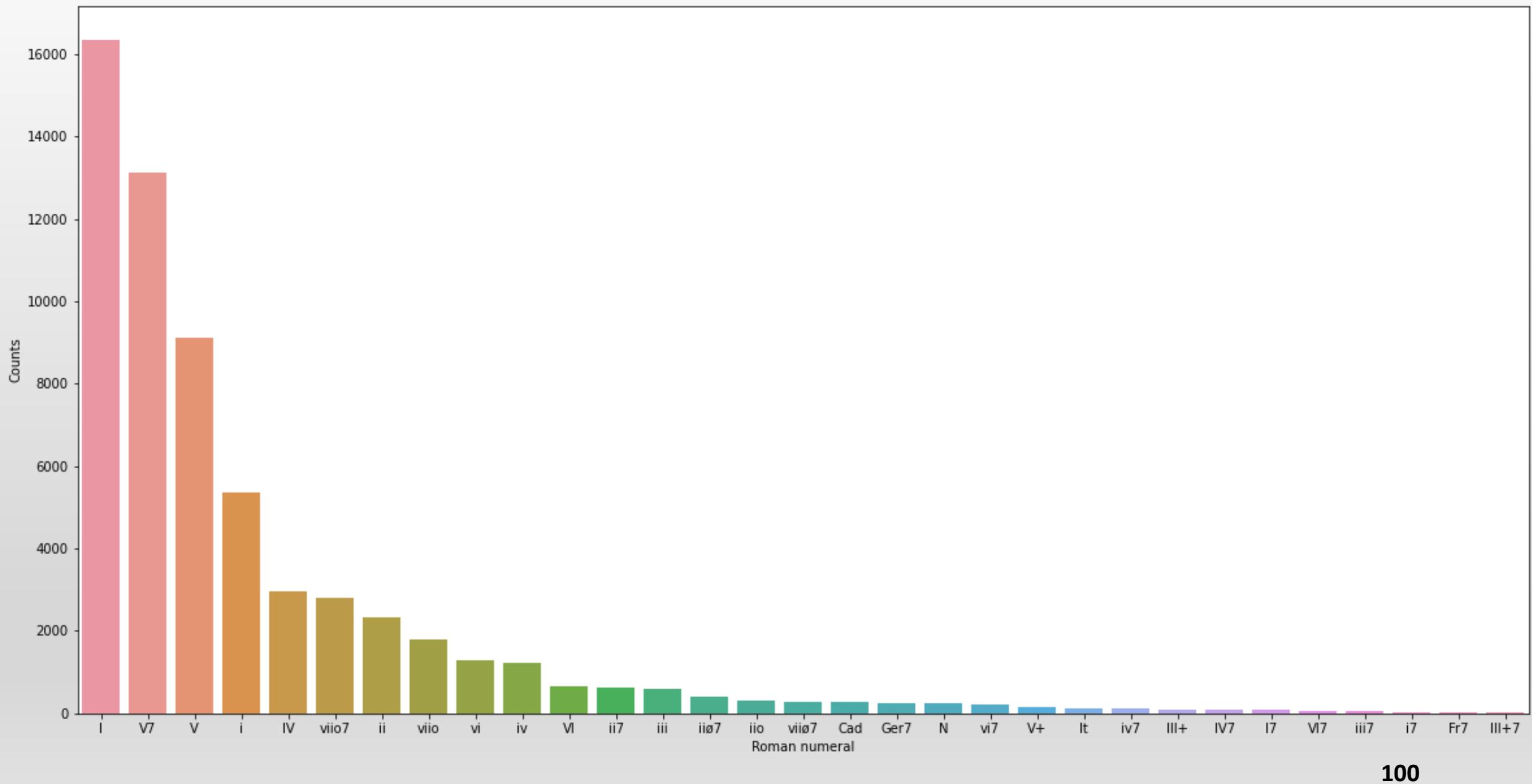
AugmentedNet

21 out of 31 classes

Unrecognized classes

3 out of 31

Numerator	Occurrence (%)	Melisma	C&S21	Mi21	M&R21	AugmentedNet
III⁺⁷	0.02	0	0	0	0	0
Fr⁷	0.03	0	0	0	0	57.9
III⁺	0.04	0	0	4.1	8.3	41.7
i⁷	0.06	19.0	0	0	23.2	0
I⁷	0.07	0	0	0	0	0
V⁺	0.09	0	12.8	10.6	23.4	40.4
IV⁷	0.09	0	0	0	0	0
iii⁷	0.10	5.4	22.2	0	0	3.7
iv⁷	0.16	33.3	7	20.0	22.2	37.8
vi⁷	0.19	3.2	29.3	1.0	0	15.6
VI⁷	0.19	0	0	0	7.7	15.3
It	0.22	0	24.3	0	0	32.4
vii^{o7}	0.25	0	7.3	0	1.5	14.3
Ger⁷	0.43	0	0	25.7	0.9	44.6
ii^o	0.47	0.6	14.6	26.0	13.5	35.0
N	0.51	0	43.4	64.2	35.0	32.0
ii^{o7}	0.69	19.1	14.4	6.6	34.2	41.5
iii	0.78	18.7	2.0	27.0	27.8	23.2
ii⁷	0.91	27.8	37.0	19.1	17.0	41.0
Cad⁶₄	1.13	0	0	0	0	6.9
VI	1.40	5.5	1.6	25.0	36.0	47.4
vi	1.82	21.6	28.5	31.2	28.2	35.2
iv	1.88	28.1	29	56.6	37.2	51.4
vii^o	2.37	5.3	21.9	11.4	25.2	43.8
ii	3.70	27.4	45.5	54.9	33.0	60.3
IV	3.78	28.0	50.5	58.9	26.7	57.0
vii^{o7}	5.46	0	50.7	56.8	47.0	71.0
i	10.93	49.4	71.6	78.8	56.0	82.3
V	13.89	37.6	44.9	51.9	56.3	60.9
V⁷	23.17	43.1	57.9	68.0	55.9	68.6
I	25.18	60.5	75.7	83.1	59.5	84.3



100

References

Nápoles López, Néstor. 2022. “Automatic Roman Numeral Analysis in Symbolic Music Representations.” PhD Thesis, McGill University.
<https://escholarship.mcgill.ca/concern/theses/qr46r6307>.

Nápoles López, Néstor, Mark Gotham, and Ichiro Fujinaga. 2021. “AugmentedNet: A Roman Numeral Analysis Network with Synthetic Training Examples and Additional Tonal Tasks.” In *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, 404–11.

End-user application of the
model

Chord Autocompletion

- Sibelius, June 2023 Release
 - <https://www.avid.com/resource-center/whats-new-in-sibelius-june-2023>

Outline

- Introduction to symbolic music data
 - Representations
 - Symbolic vs Audio Data
 - Symbolic music tasks
- Concrete example
 - Automatic Roman Numeral Analysis
 - End-user applications
- Additional resources
 - Deep learning for music generation
 - Datasets
 - Libraries

Deep Learning Techniques for Music Generation

- <https://arxiv.org/abs/1709.01620>
- Nice literature review of (early) symbolic music generation
- Recommend Chapter 4 on “Representation”
- Approaches covered from simple feedforward networks to (older) transformer models
 - MiniBach implementation: <https://github.com/napulen/MiniBach>
 - CoCoNet (Bach Doodle): <https://www.google.com/doodles/celebrating-johann-sebastian-bach>
 - DeepBach model: <https://github.com/Ghadjeres/DeepBach>

Datasets and corpora

- Bach Chorales
 - <https://github.com/craigsapp/bach-370-chorales>
- Classical music (Humdrum format)
 - <http://kern.ccarh.org/>
- MIDI data
 - Lakh dataset: <https://colinraffel.com/projects/lmd/>
 - Hook Theory Lead Sheet Dataset: <https://github.com/wayne391/lead-sheet-dataset>

Software Libraries

- Processing MusicXML/Humdrum/MIDI data
 - MIT's music21: <http://web.mit.edu/music21/>
- Processing MIDI and piano-roll data
 - pretty_midi: <https://craffel.github.io/pretty-midi/>
 - mido: <https://github.com/mido/mido/>

Thank you