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

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

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

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

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

Prélude in C Minor

Op. 28 No. 20

Frédéric François Chopin (1810–1849)

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*

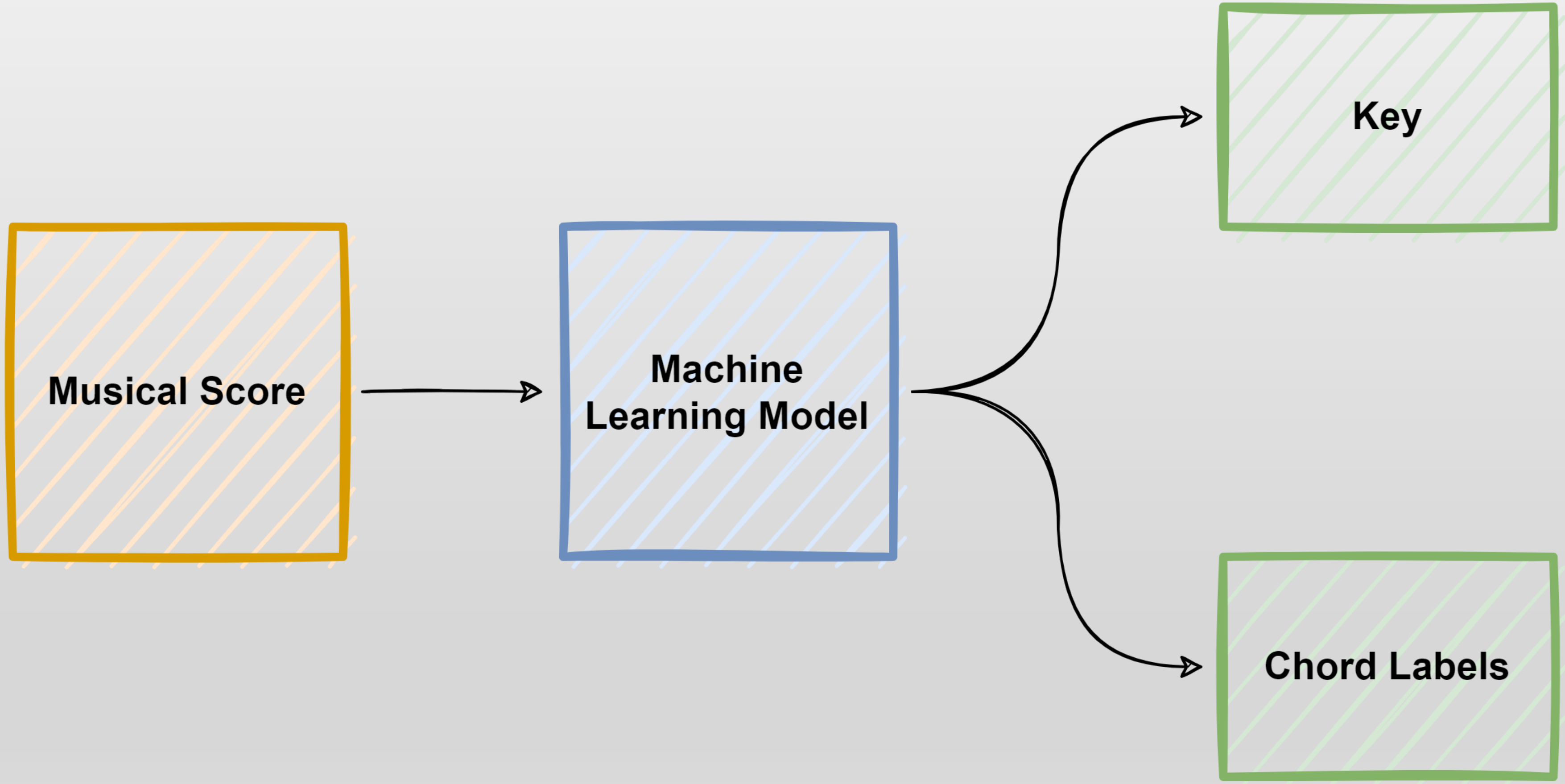
Automatic Roman Numeral Analysis

- 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*

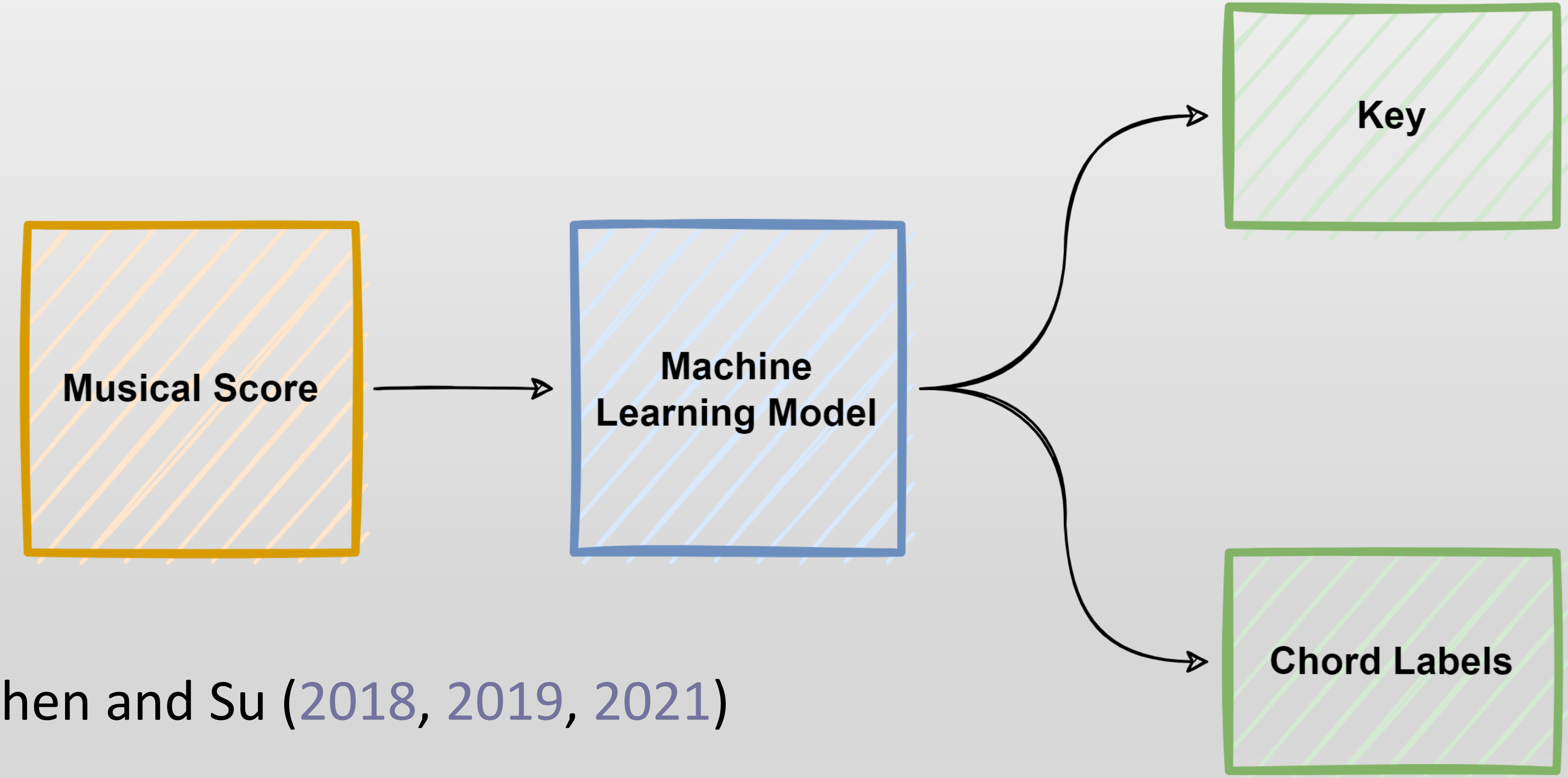
Automatic Roman Numeral Analysis

- 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

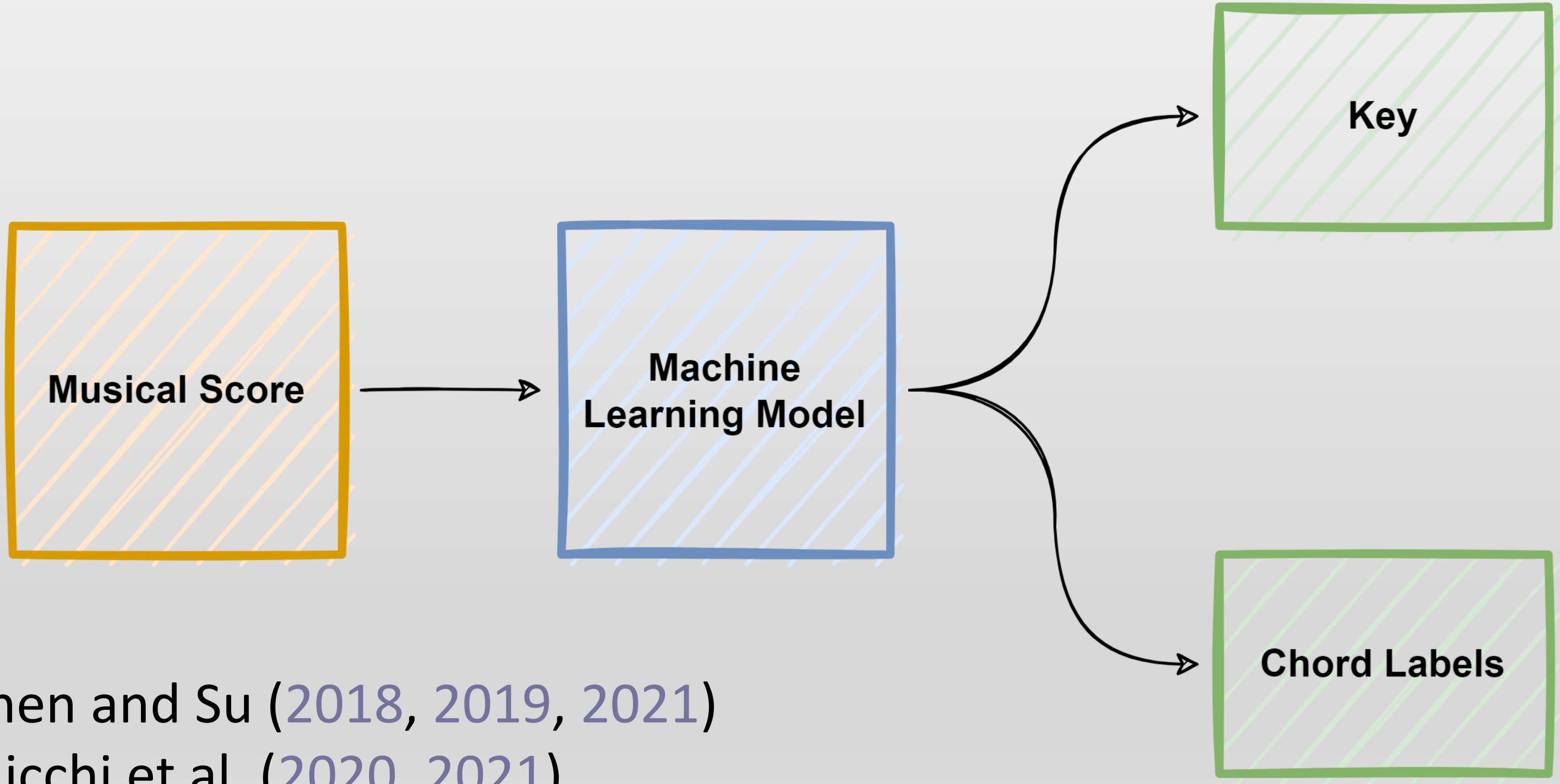


Multitask Learning Layout



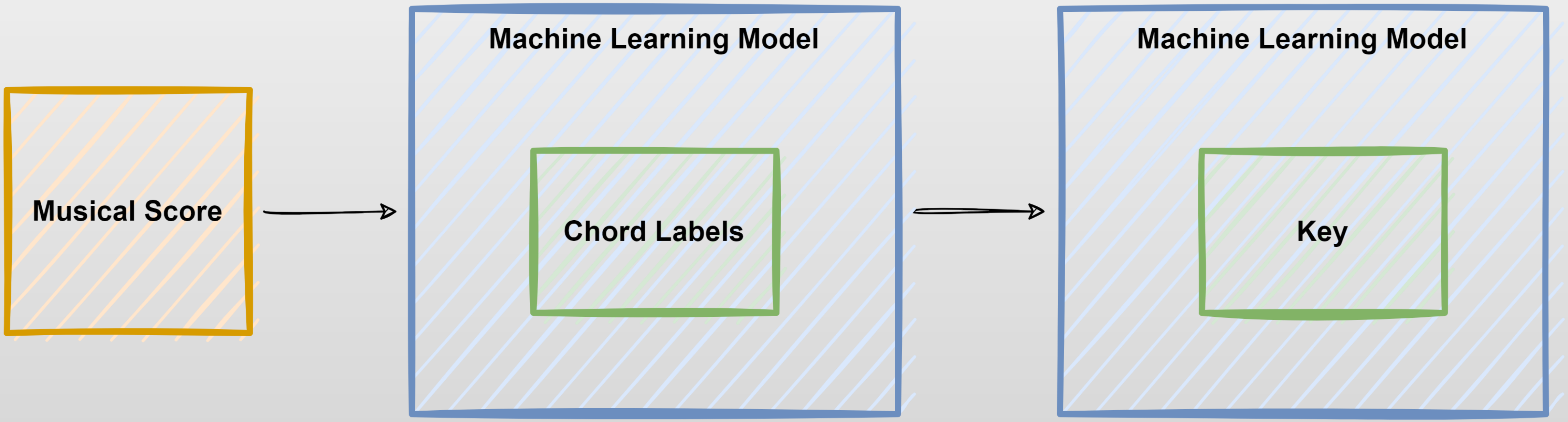
- Chen and Su (2018, 2019, 2021)

Multitask Learning Layout

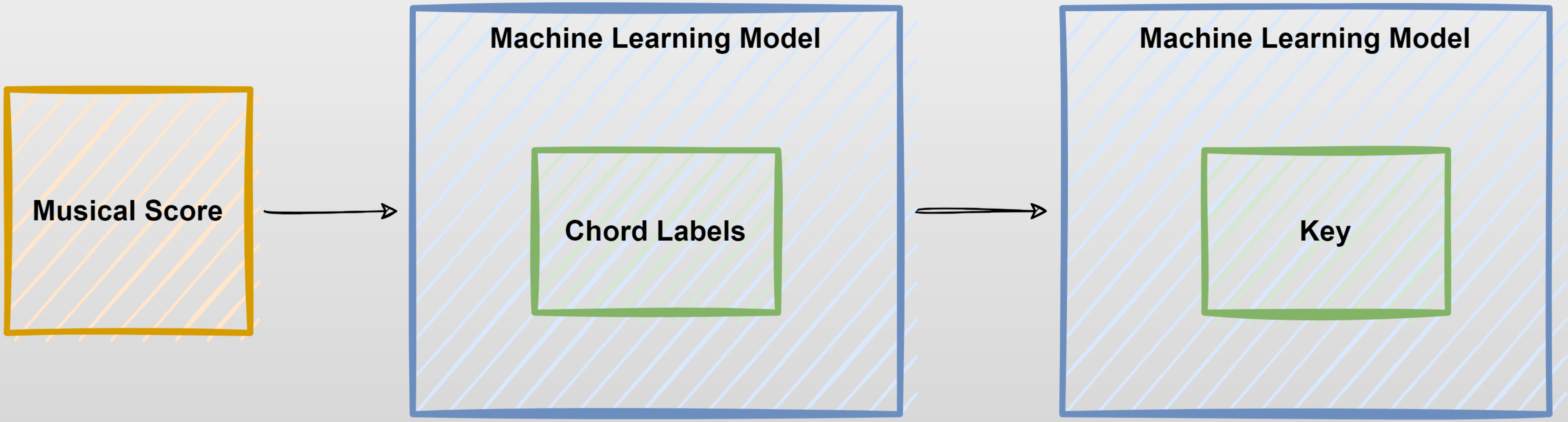


- Chen and Su (2018, 2019, 2021)
- Micchi et al. (2020, 2021)

Modular Analysis

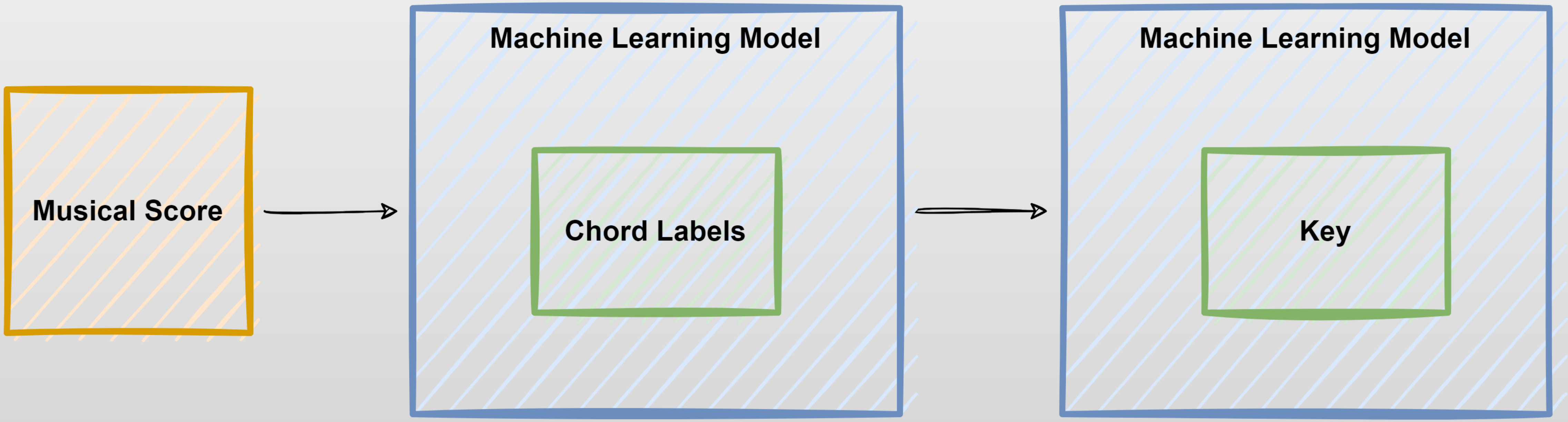


Modular Analysis



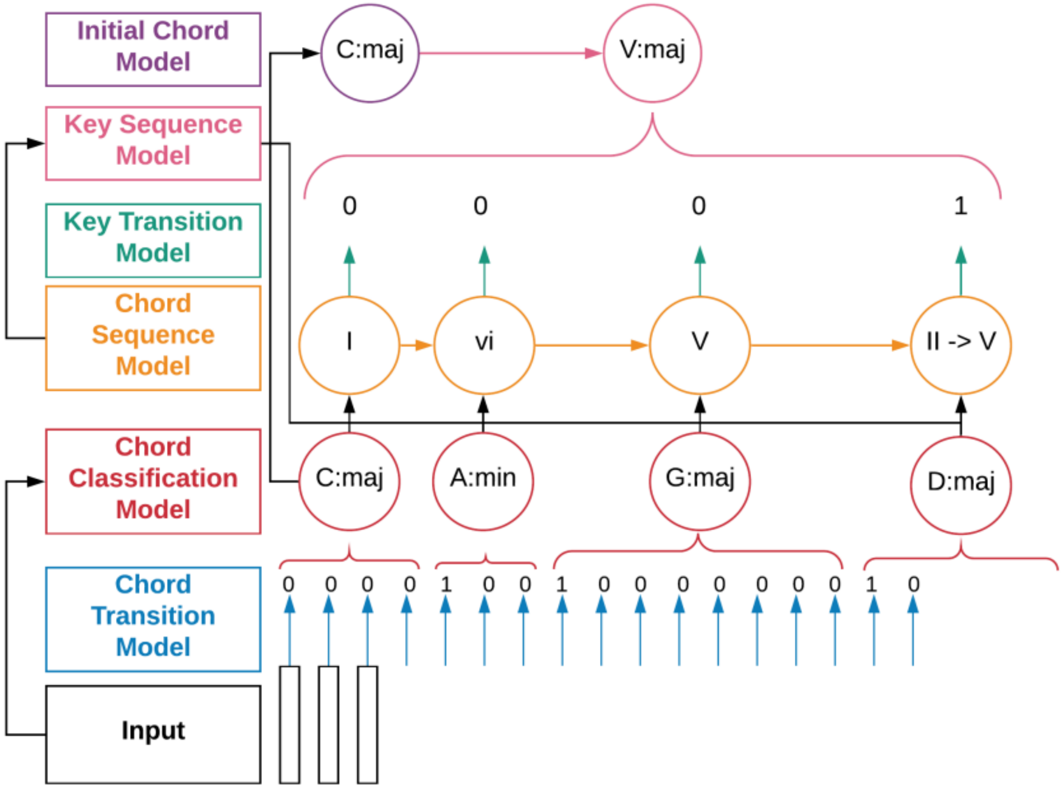
- Melisma (2003)

Modular Analysis

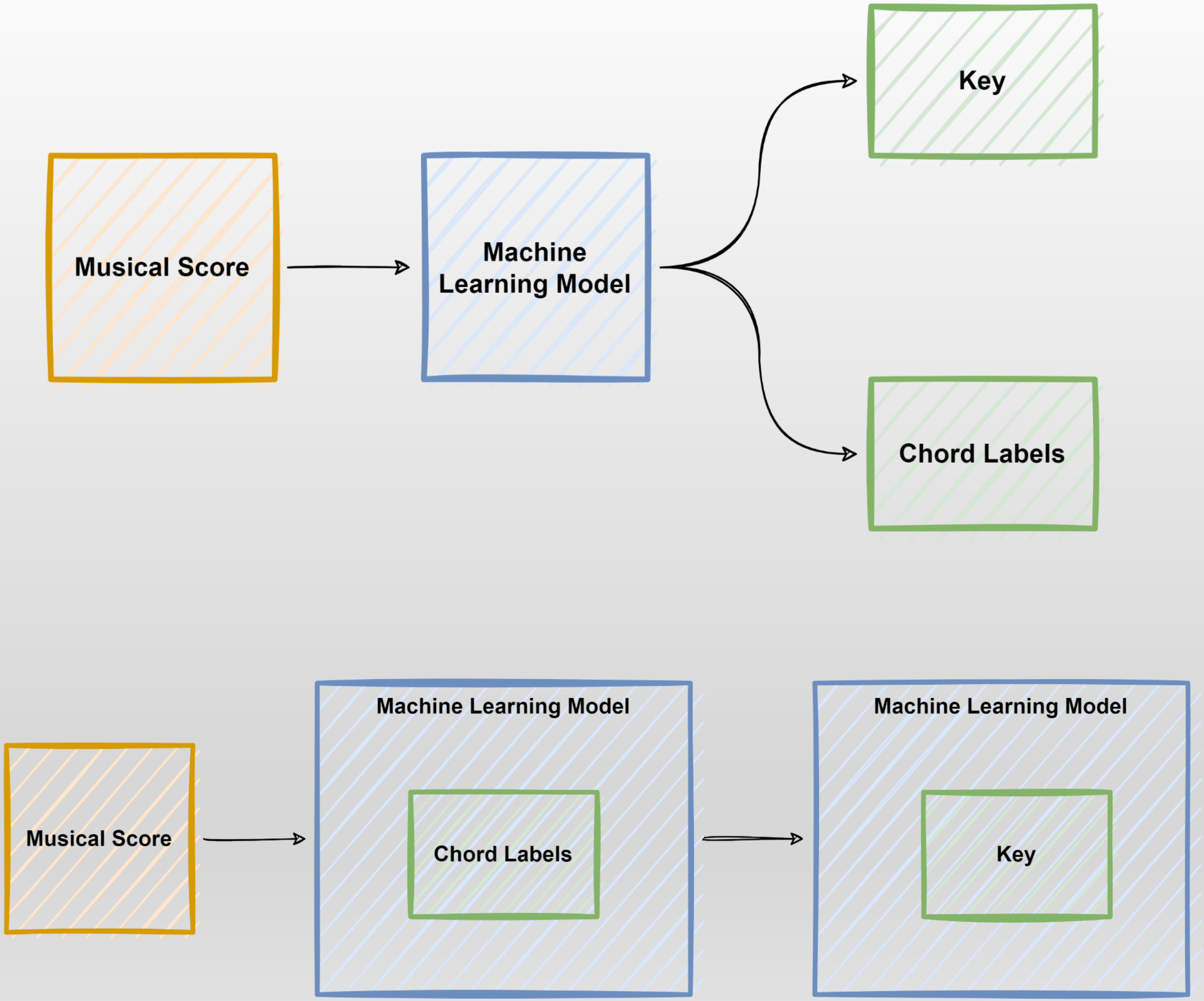


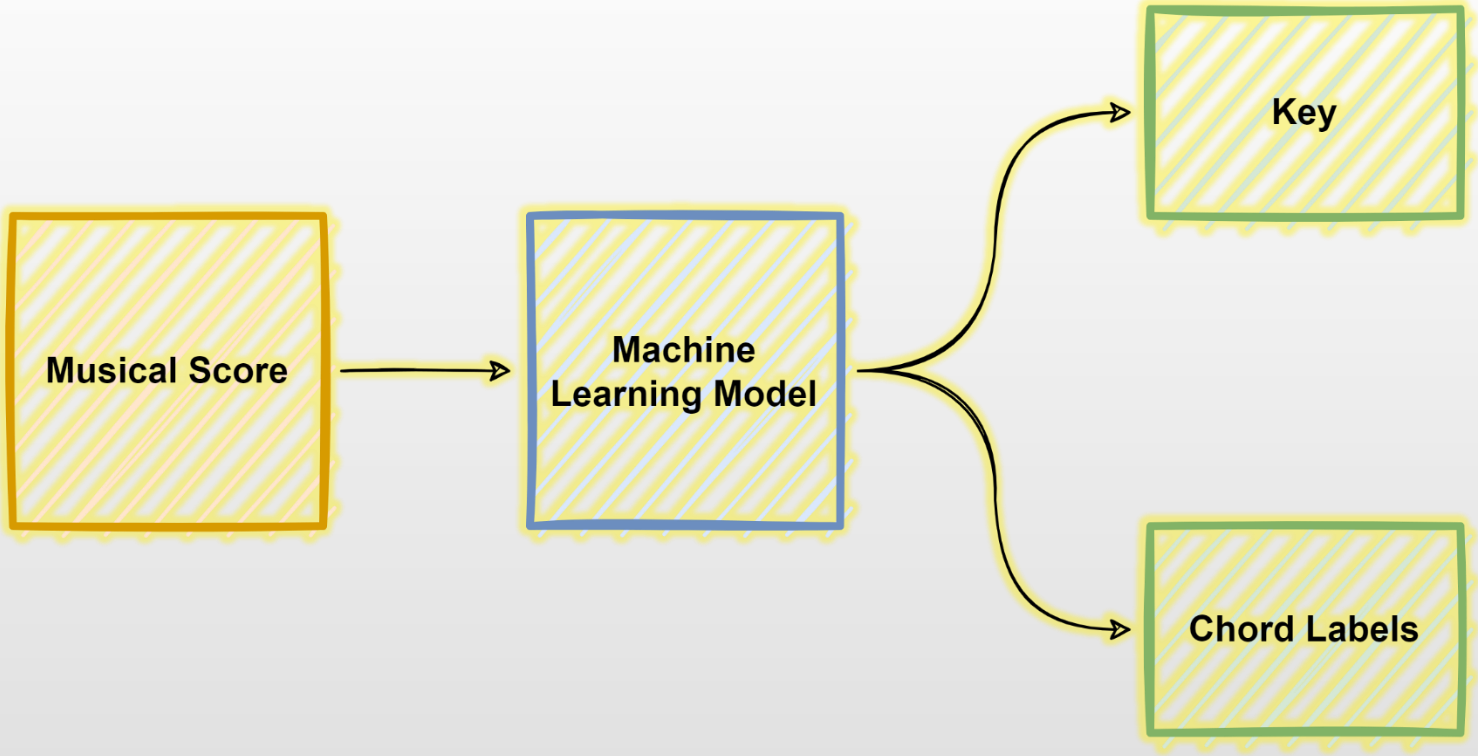
- Melisma (2003)
- McLeod and Rohrmeier (2021)

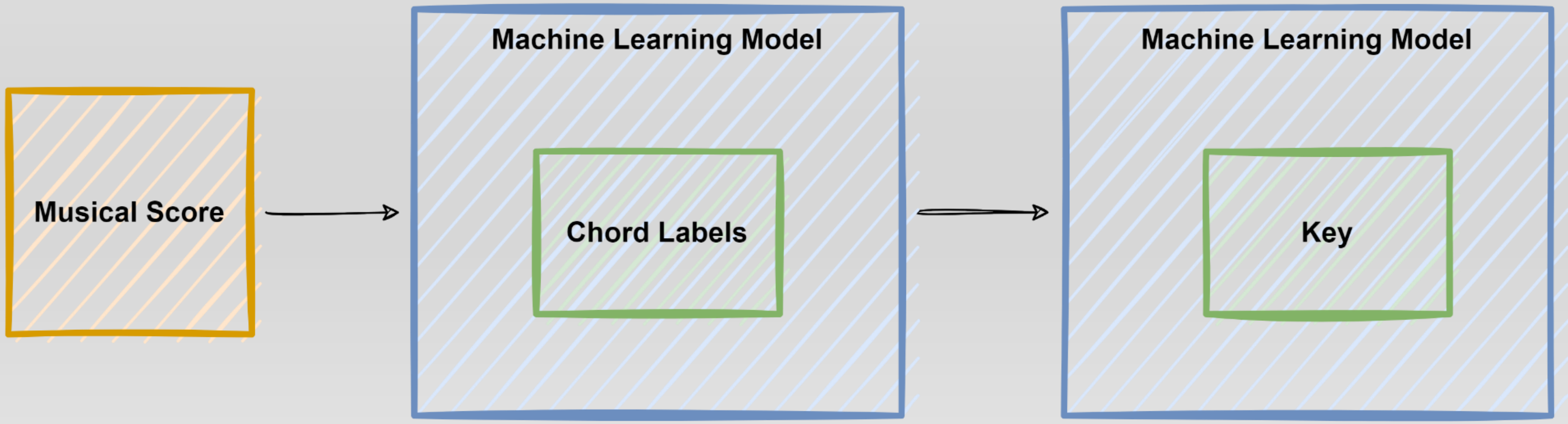
Modular Analysis



- McLeod and Rohrmeier (2021)


Musical Score
Machine Learning Model
Key
Chord Labels
Musical Score
Machine Learning Model
Chord Labels
Machine Learning Model
Key

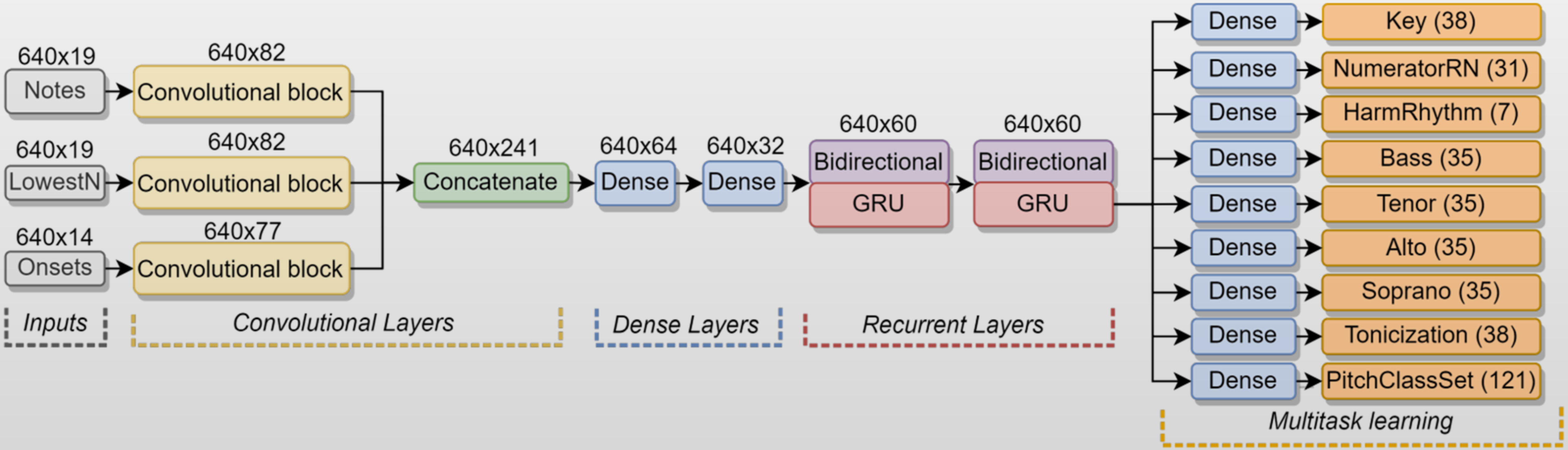

Musical Score
Machine Learning Model
Key
Chord Labels


Musical Score
Machine Learning Model
Chord Labels
Machine Learning Model
Key

Contributions of my research

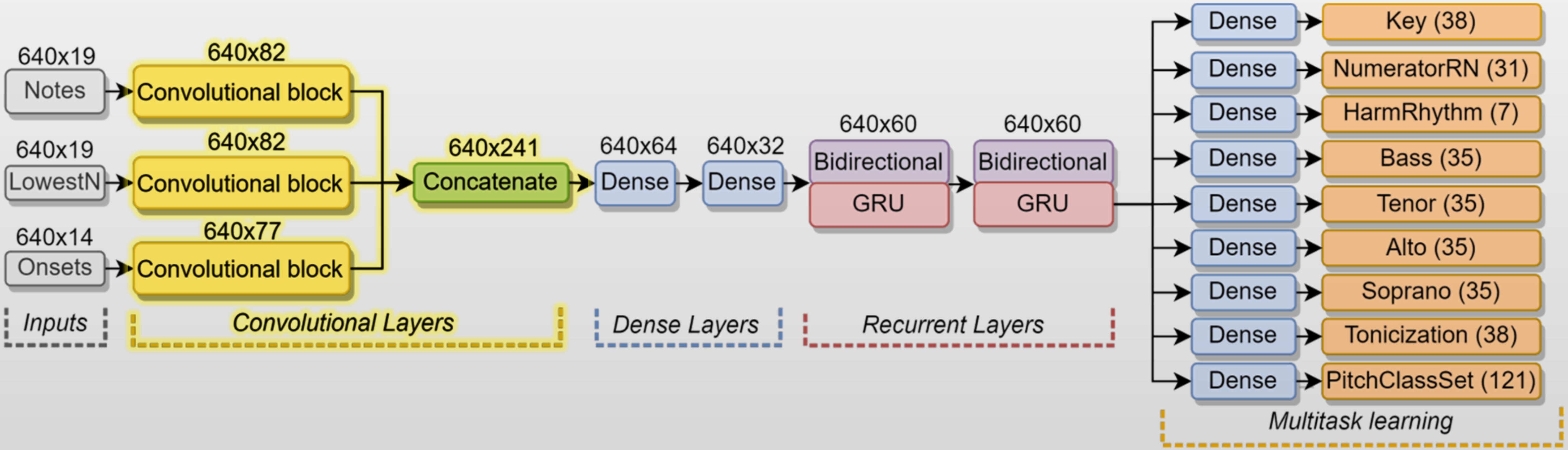
- A new neural network architecture, AugmentedNet

Neural Network Architecture



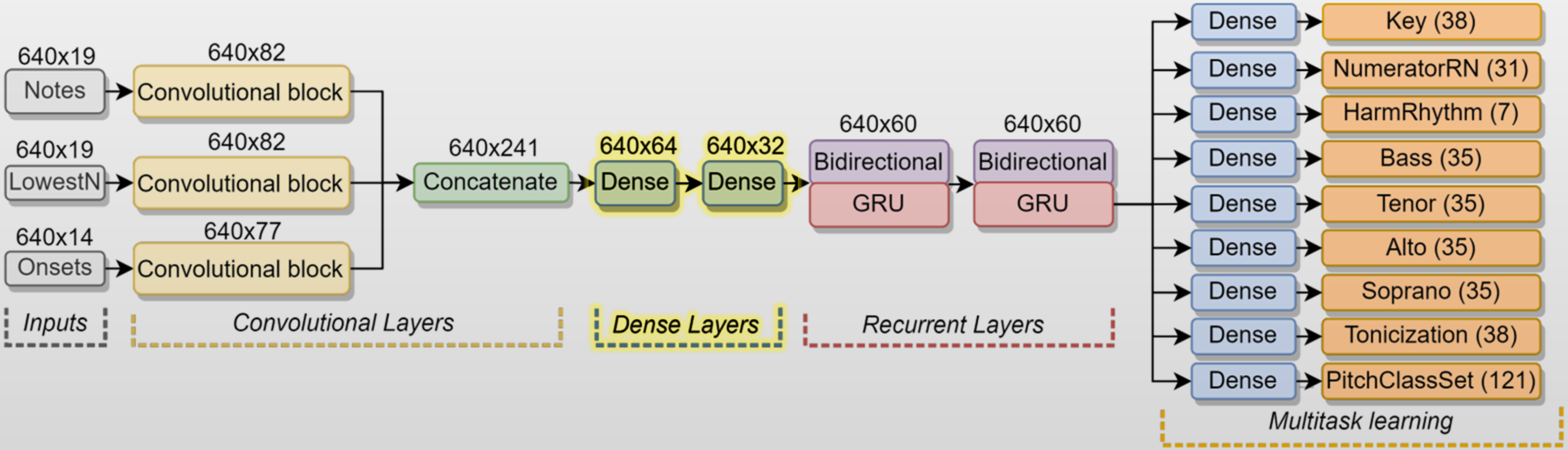
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



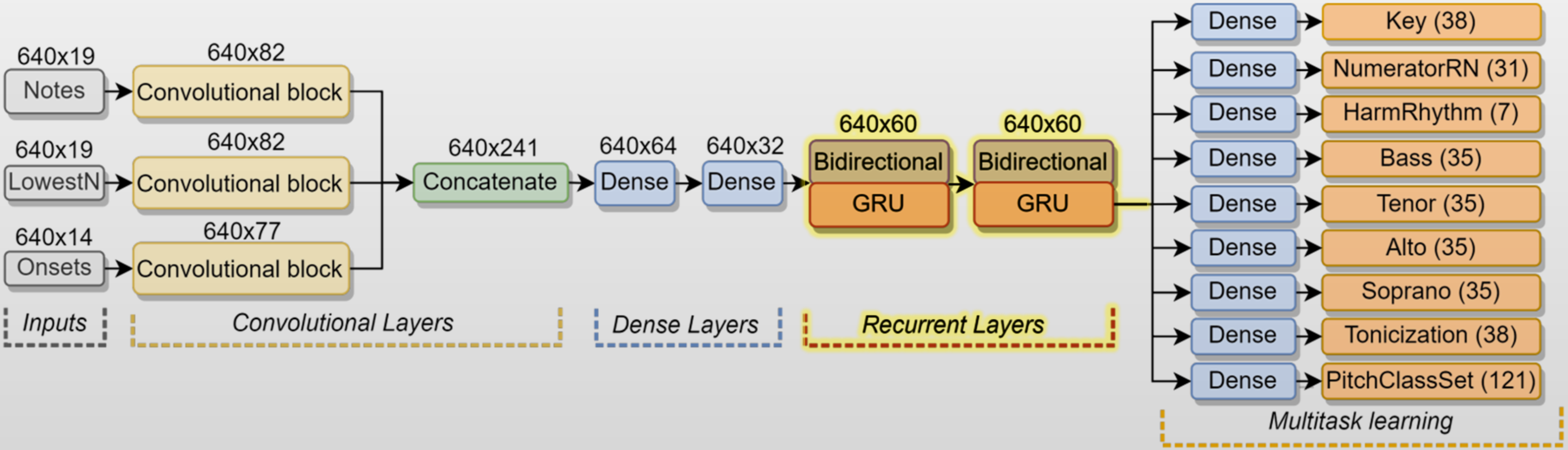
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



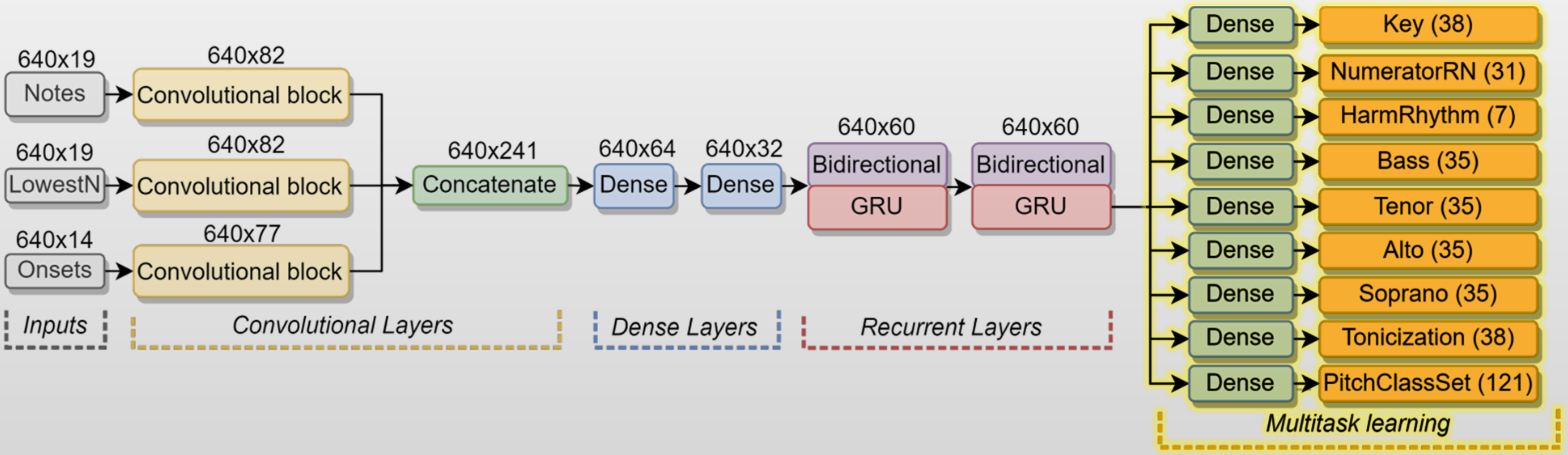
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



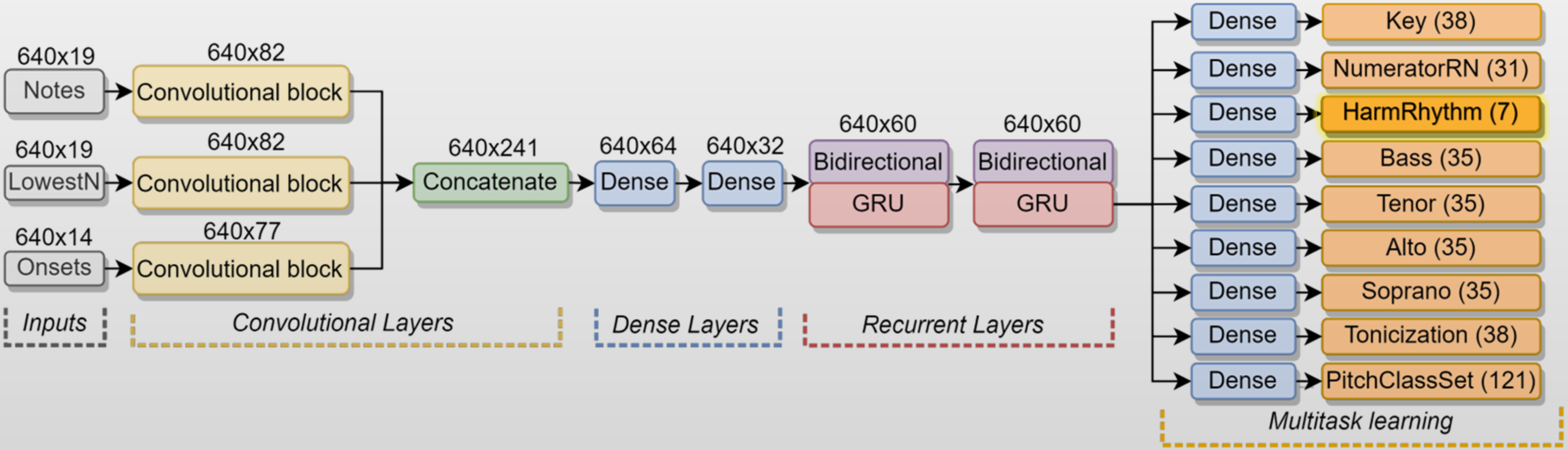
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



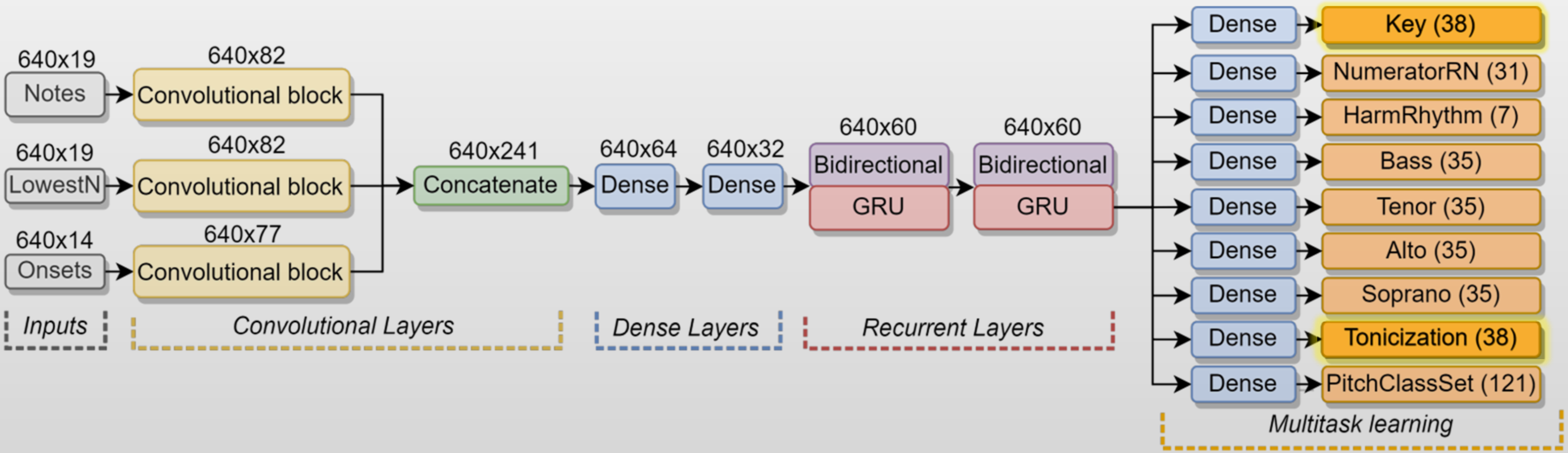
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



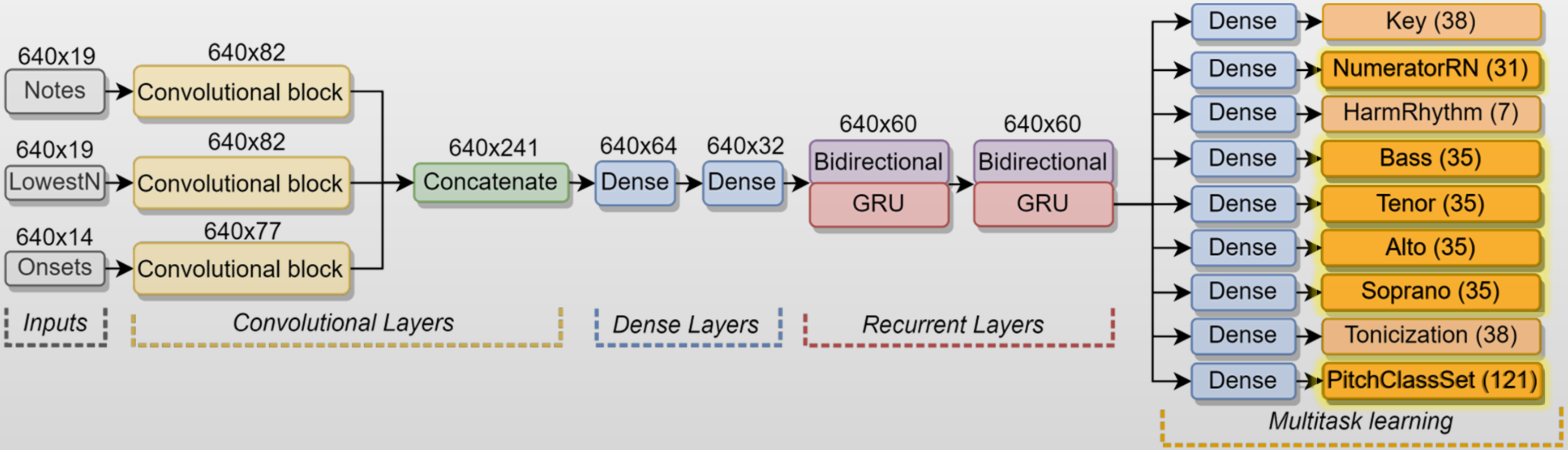
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



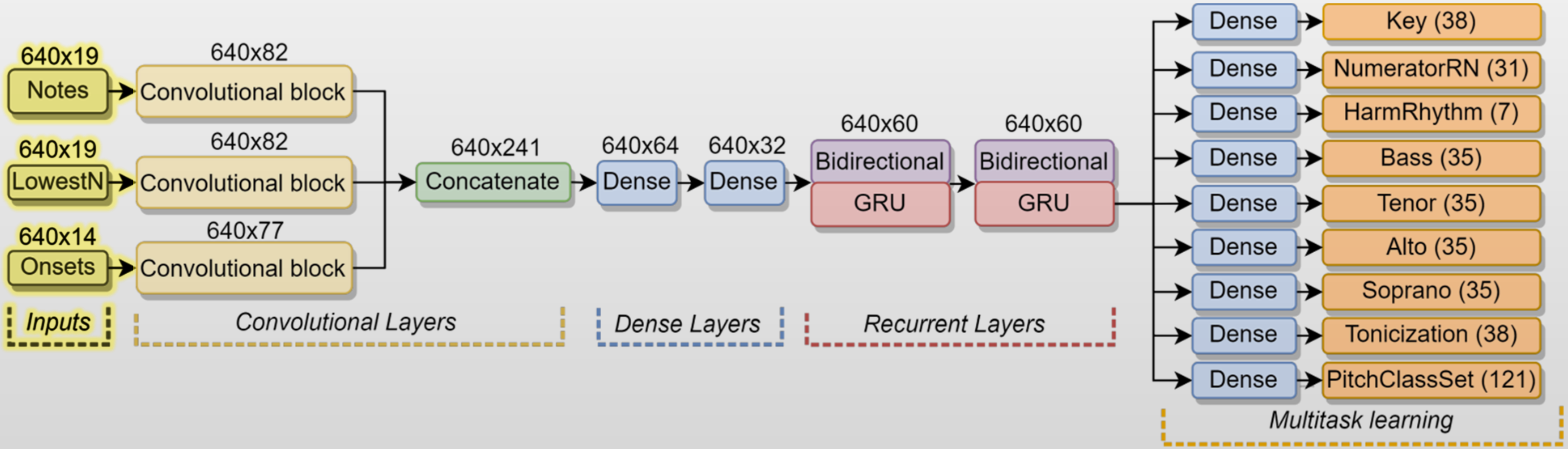
Convolutional Recurrent Neural Network (**CRNN**)

Neural Network Architecture



Convolutional Recurrent Neural Network (**CRNN**)

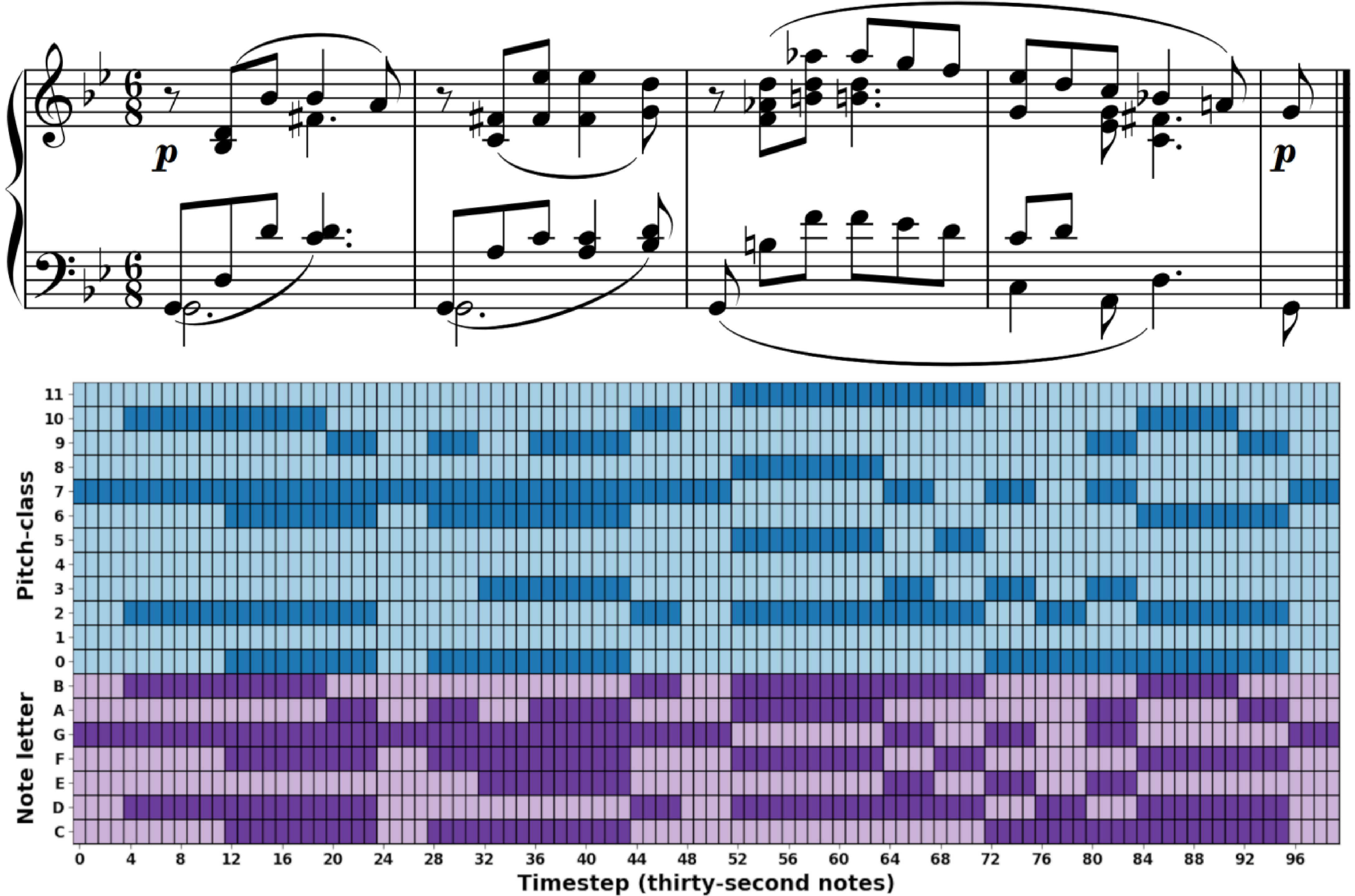
Neural Network Architecture



Convolutional Recurrent Neural Network (**CRNN**)

Input Components

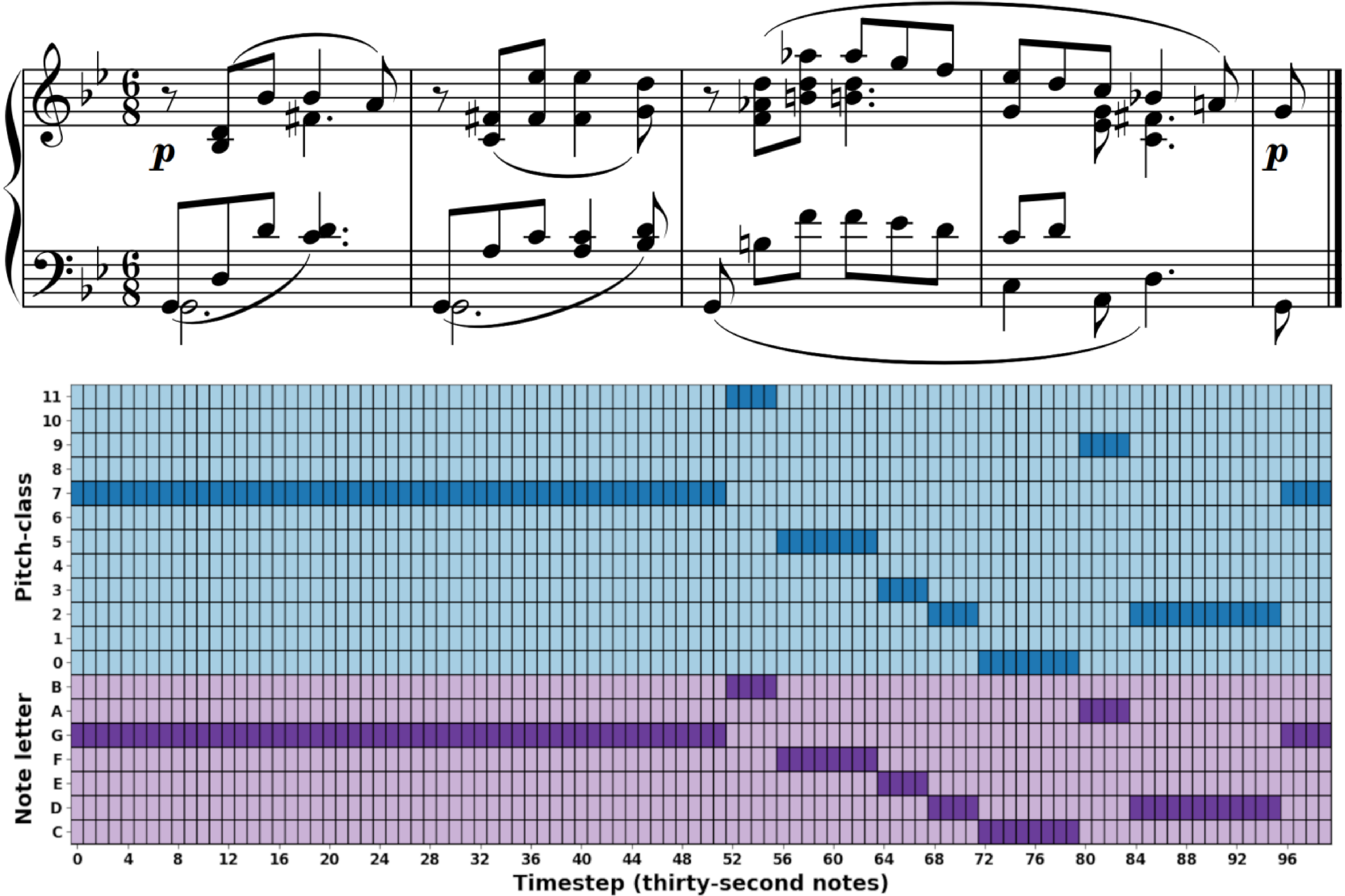
- All the sounding notes (**Notes19**)
  - All the pitches (with spelling) that are "sounding" at the current timestep
  - No octave information


p
p
Pitch-class
11
10
9
8
7
6
5
4
3
2
1
0
Note letter
B
A
G
F
E
D
C
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96
Timestep (thirty-second notes)

Input Components

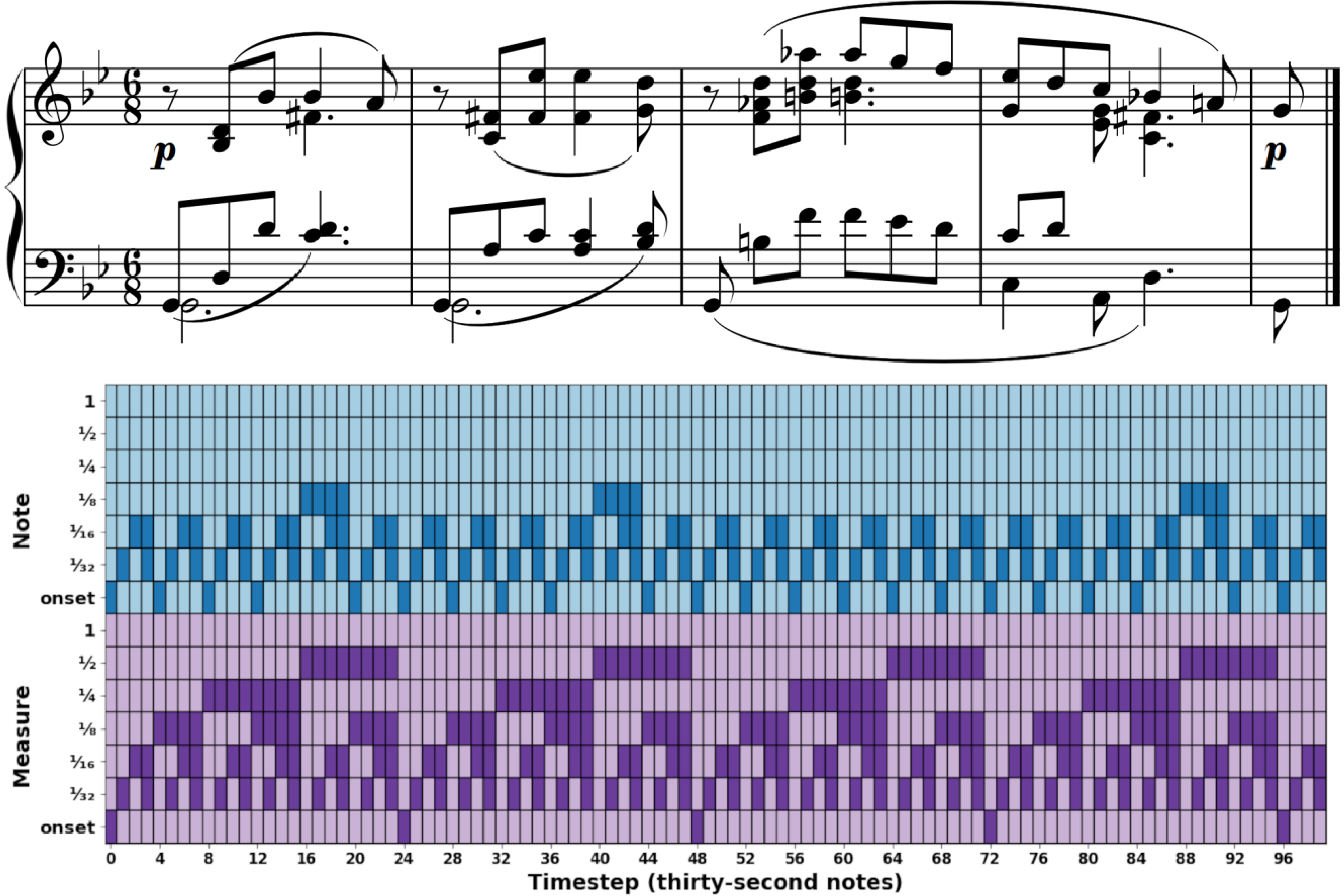
- All the sounding notes (**Notes19**)
  - All the pitches (with spelling) that are “sounding” at the current timestep
  - No octave information

- Lowest-sounding note (**LowestNote19**)
  - Similar to **Notes19,** but only the lowest-sounding note at the current timestep


p
p
Pitch-class
11
10
9
8
7
6
5
4
3
2
1
0
Note letter
B
A
G
F
E
D
C
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96
Timestep (thirty-second 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
- Onsets (**Onsets14**)
  - The location of note onsets (attacks) as well as the location of changes of measure


p
p
Note
1
½
¼
⅛
1⁄16
1⁄32
onset
Measure
1
½
¼
⅛
1⁄16
1⁄32
onset
0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96
Timestep (thirty-second notes)

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

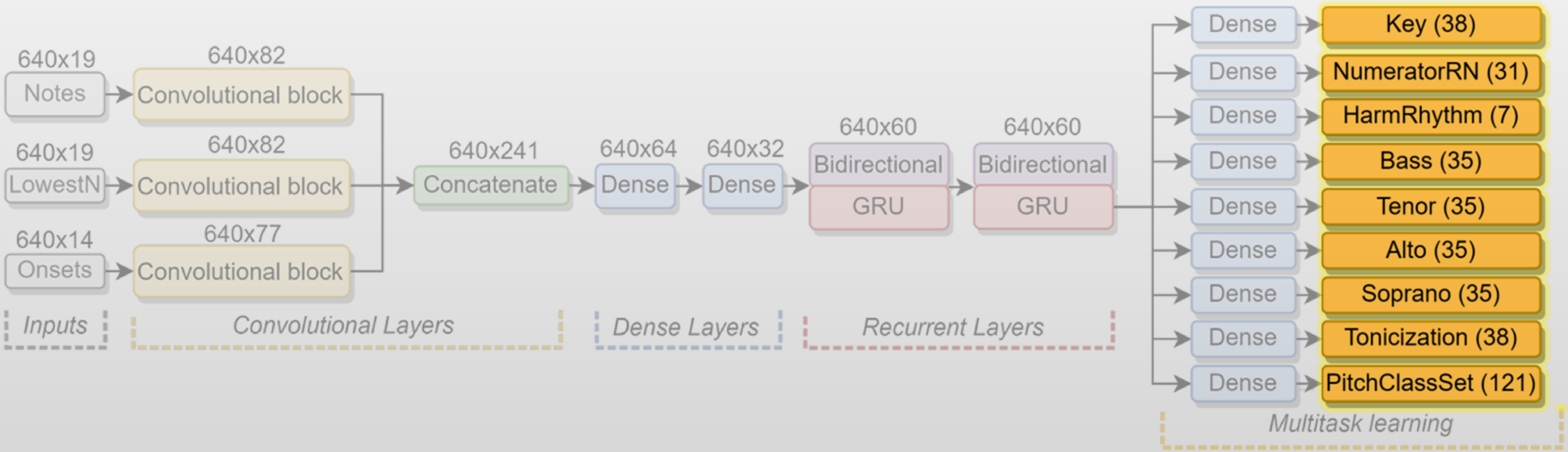
Average accuracy performance in a 5-fold cross-validation experiment

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

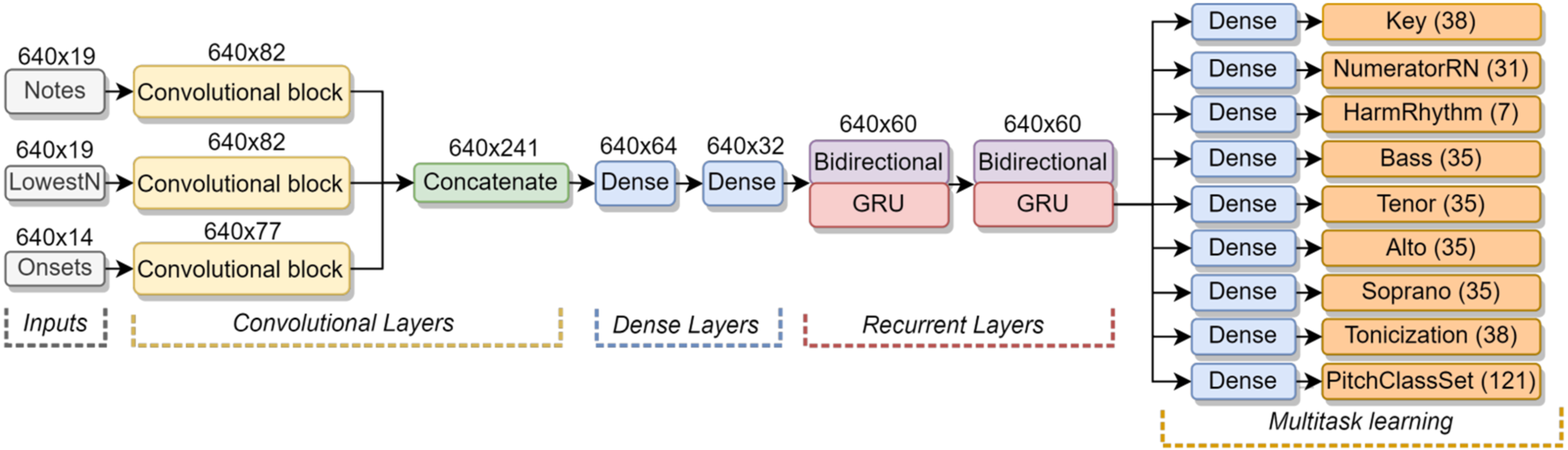
Average accuracy performance in a **5-fold cross-validation** experiment

Ablation Studies

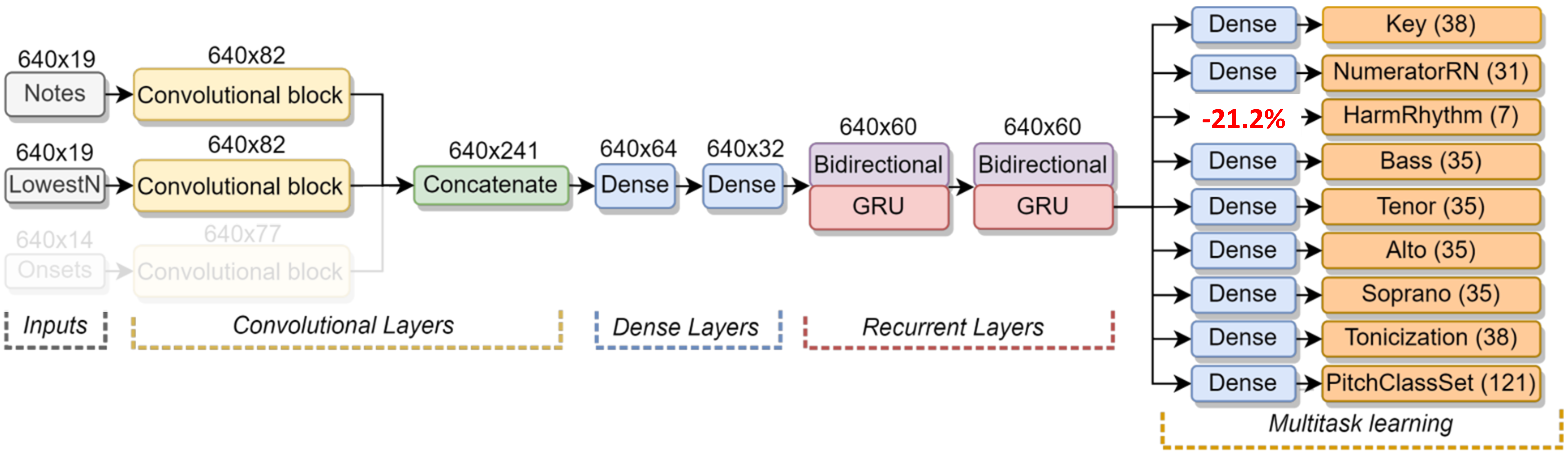
Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	$68.1_{2.2}$	$74.5_{2.2}$	$78.9_{0.9}$	$78.8_{3.4}$	$73.5_{1.9}$	$61.3_{2.8}$	$72.3_{1.7}$	$71.3_{2.5}$	$80_{1.8}$
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No **LowestNote19**	**-13.1**	**-16.5**	0.0	+0.3	+0.3	+0.4	-5.0	**-15.4**	+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

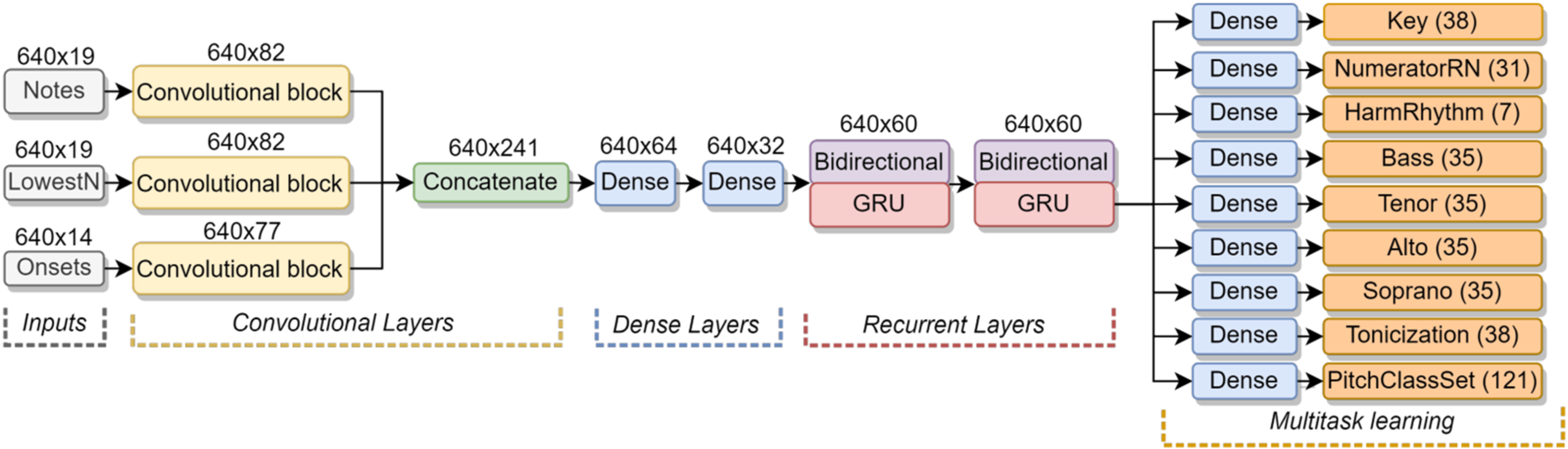
Average accuracy performance in a **5-fold cross-validation** experiment

Ablation Studies

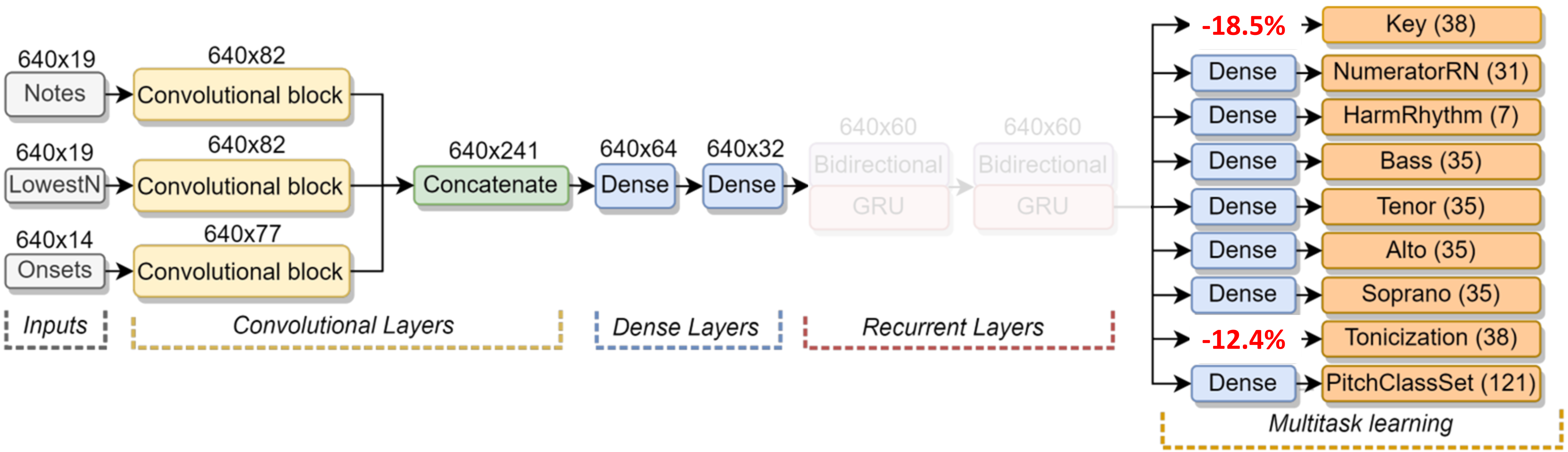
Model	A35	B35	HR7	K38	PCS121	N31	S35	T35	KT38
Baseline (μ_σ)	$68.1_{2.2}$	$74.5_{2.2}$	$78.9_{0.9}$	$78.8_{3.4}$	$73.5_{1.9}$	$61.3_{2.8}$	$72.3_{1.7}$	$71.3_{2.5}$	$80_{1.8}$
Pitch spelling	0.0	-0.2	+0.2	-0.1	+0.1	+0.1	+0.1	-0.1	+0.1
No **LowestNote19**	**-13.1**	**-16.5**	0.0	+0.3	+0.3	+0.4	-5.0	**-15.4**	+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	Published on
1. Annotated Beethoven Corpus (**ABC**)	Neuwirth et al. (2018)
2. Beethoven Piano Sonatas (**BPS**)	Chen and Su (2018)
3. Haydn “Sun” String Quartets (**HaydnSun**)	Nápoles López (2017)
4. Key Modulations and Tonicizations Dataset (**KMT**)	Nápoles López et al. (2020)
5. Mozart Piano Sonatas (**MPS**)	Hentschel et al. (2021)
6. Theme and Variations Encodings with Roman numerals (**TAVERN**)	Devaney et al. (2015)
7. When in Rome (**WiR**)	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

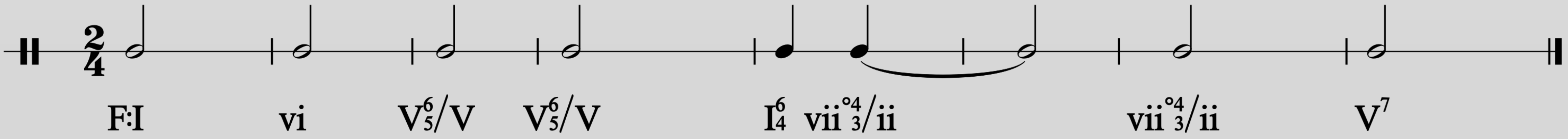
Data Augmentation - Transposition

- 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**

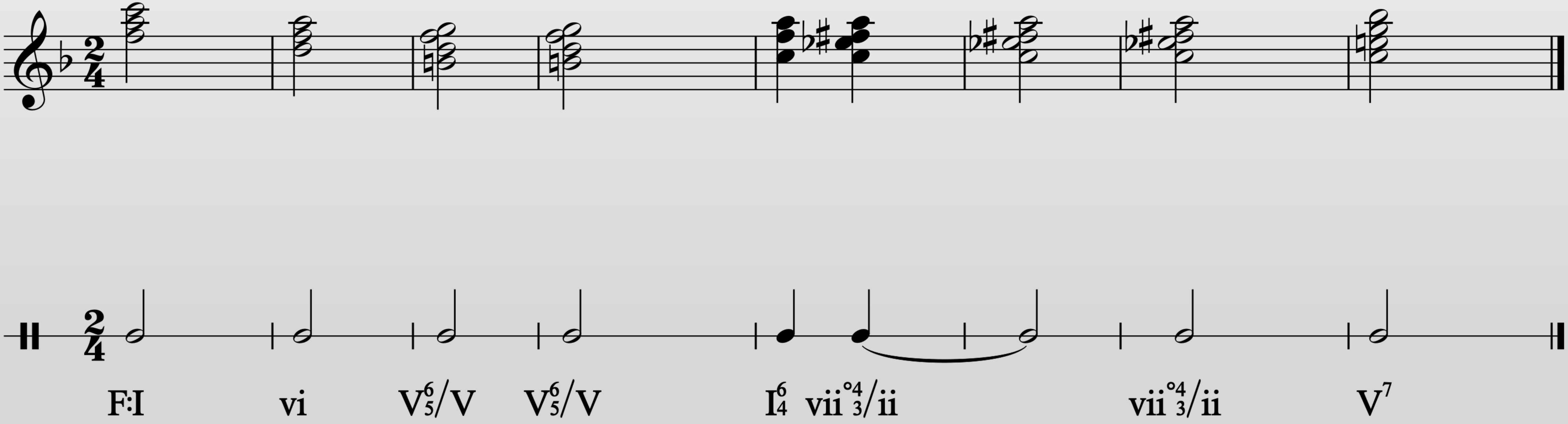
Contributions of my research

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

Data Augmentation by Synthesis

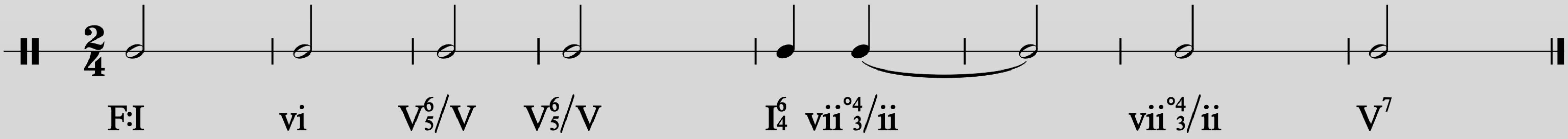


Data Augmentation by Synthesis

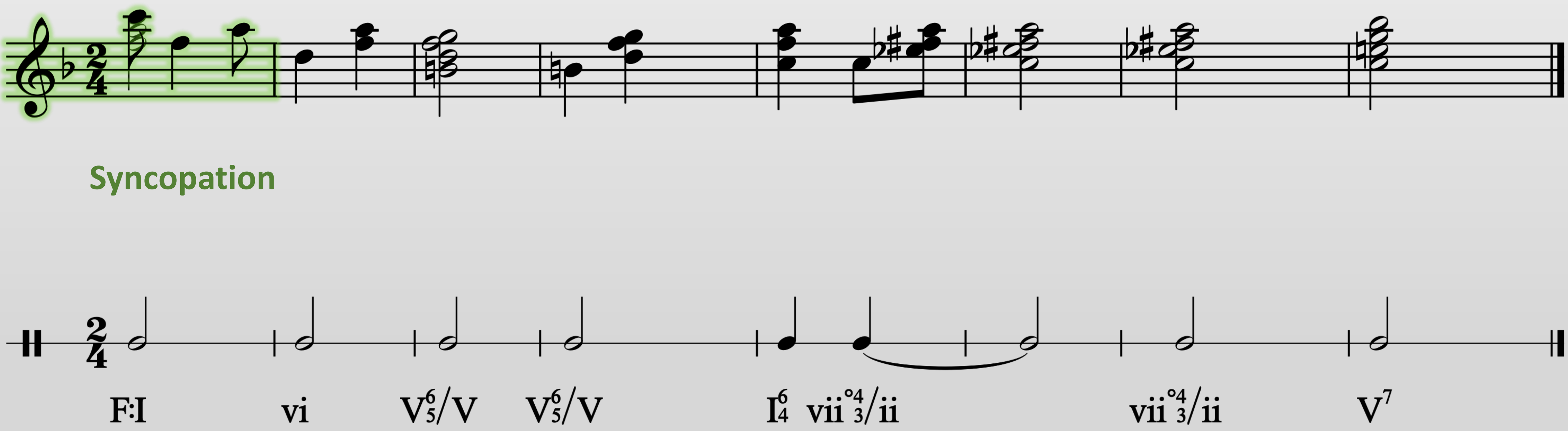


Data Augmentation by Synthesis

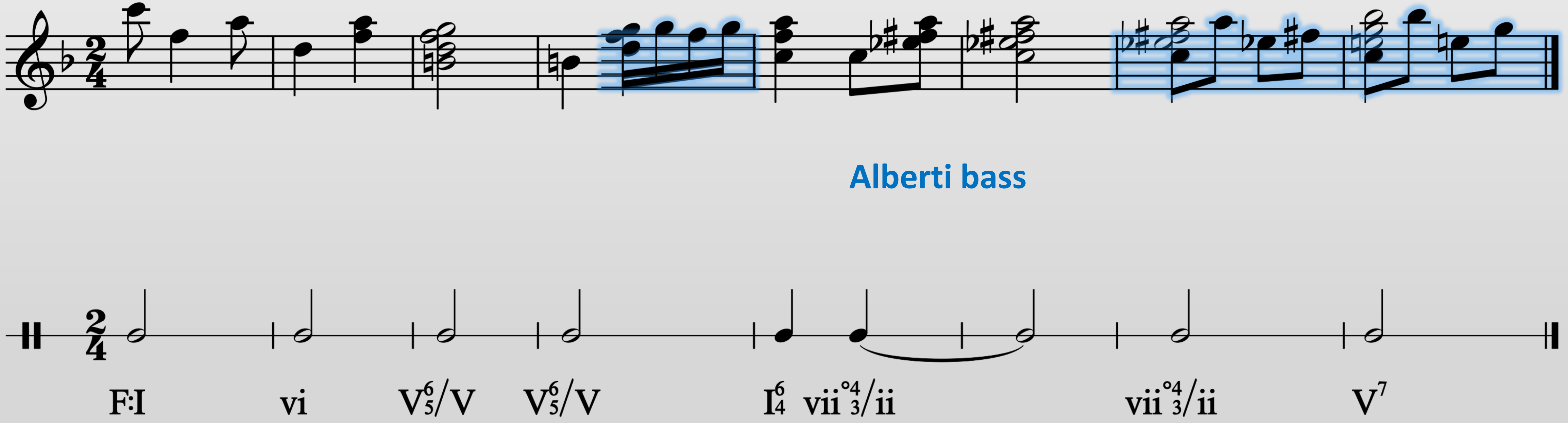
Separate the bass



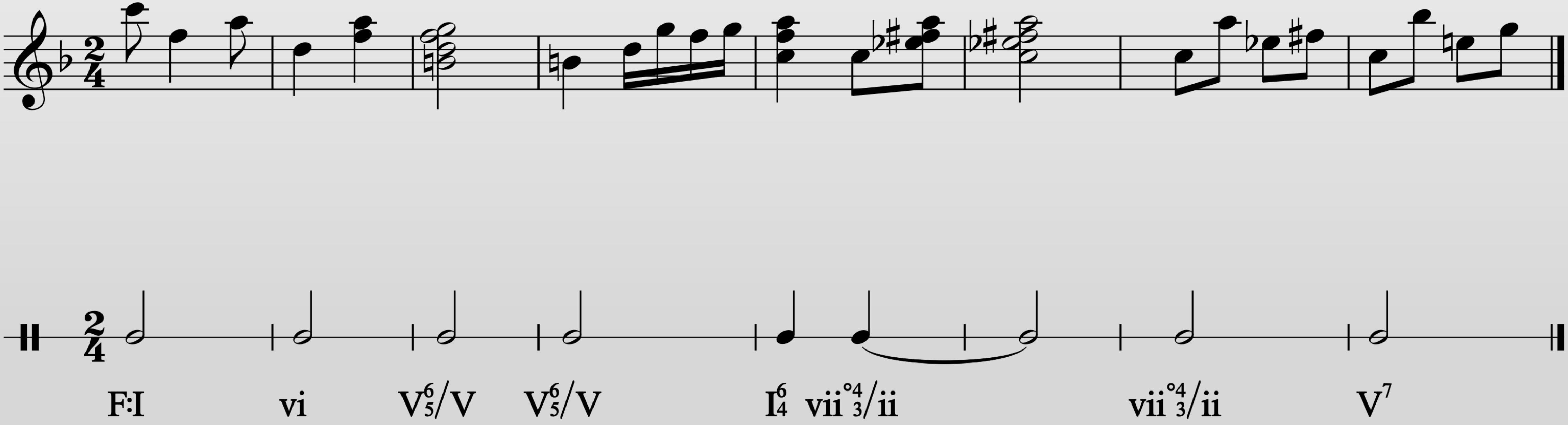
Data Augmentation by Synthesis



Data Augmentation by Synthesis



Data Augmentation by Synthesis



Effects of Data Augmentation

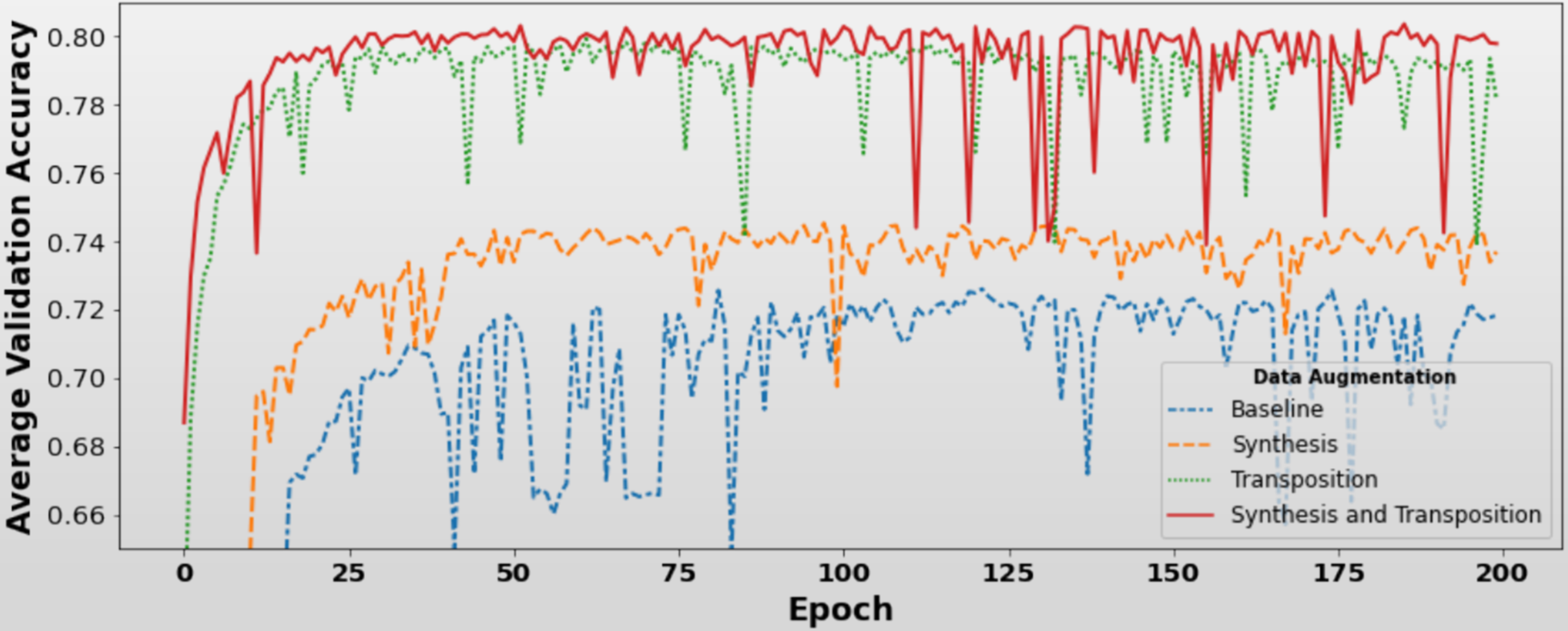


Figure 6.14: *Average validation accuracy achieved in the four experiments with data augmentation. The average accuracy value is across the seven datasets at the given epoch.*

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	$65.2_{7.04}$	$67.2_{7.00}$	$70.7_{6.09}$	$\mathbf{72.2}_{5.71}$
BPS	$62.5_{9.25}$	$64.6_{6.62}$	$74.7_{5.89}$	$\mathbf{75.7}_{5.70}$
HaydnSun	$54.4_{11.76}$	$59.4_{9.64}$	$68.2_{8.61}$	$\mathbf{68.7}_{6.87}$
KMT	$78.8_{6.02}$	$81.6_{6.65}$	$85.3_{5.84}$	$\mathbf{88.7}_{3.60}$
MPS	$73.2_{7.52}$	$75.2_{6.86}$	$\mathbf{81.8}_{5.90}$	$81.5_{5.89}$
TAVERN	$69.1_{7.55}$	$71.3_{7.66}$	$76.8_{6.07}$	$\mathbf{77.0}_{6.44}$
WiR	$69.9_{8.79}$	$71.7_{7.88}$	$78.1_{5.40}$	$\mathbf{78.6}_{4.74}$

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	$65.2_{7.04}$	$67.2_{7.00}$	$70.7_{6.09}$	$\mathbf{72.2}_{5.71}$
BPS	$62.5_{9.25}$	$64.6_{6.62}$	$74.7_{5.89}$	$\mathbf{75.7}_{5.70}$
HaydnSun	$54.4_{11.76}$	$59.4_{9.64}$	$68.2_{8.61}$	$\mathbf{68.7}_{6.87}$
KMT	$78.8_{6.02}$	$81.6_{6.65}$	$85.3_{5.84}$	$\mathbf{88.7}_{3.60}$
MPS	$73.2_{7.52}$	$75.2_{6.86}$	$\mathbf{81.8}_{5.90}$	$81.5_{5.89}$
TAVERN	$69.1_{7.55}$	$71.3_{7.66}$	$76.8_{6.07}$	$\mathbf{77.0}_{6.44}$
WiR	$69.9_{8.79}$	$71.7_{7.88}$	$78.1_{5.40}$	$\mathbf{78.6}_{4.74}$

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
Fr^{7}	0.03	0	0	0	0	**57.9**
III^{+}	0.04	0	0	4.1	8.3	**41.7**
i^{7}	0.06	19.0	0	0	**23.2**	0
I^{7}	0.07	0	0	0	0	0
V^{+}	0.09	0	12.8	10.6	23.4	**40.4**
IV^{7}	0.09	0	0	0	0	0
iii^{7}	0.10	5.4	**22.2**	0	0	3.7
iv^{7}	0.16	33.3	7	20.0	22.2	**37.8**
vi^{7}	0.19	3.2	**29.3**	1.0	0	15.6
VI^{7}	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**
Ger^{7}	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
ii^{7}	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad^{6}_{4}	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**
V^{7}	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
Fr^7	0.03	0	0	0	0	**57.9**
III^+	0.04	0	0	4.1	8.3	**41.7**
i^7	0.06	19.0	0	0	**23.2**	0
I^7	0.07	0	0	0	0	0
V^+	0.09	0	12.8	10.6	23.4	**40.4**
IV^7	0.09	0	0	0	0	0
iii^7	0.10	5.4	**22.2**	0	0	3.7
iv^7	0.16	33.3	7	20.0	22.2	**37.8**
vi^7	0.19	3.2	**29.3**	1.0	0	15.6
VI^7	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**
Ger^7	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
ii^7	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad^6_4	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**
V^7	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
Fr^{7}	0.03	0	0	0	0	**57.9**
III^{+}	0.04	0	0	4.1	8.3	**41.7**
i^{7}	0.06	19.0	0	0	**23.2**	0
I^{7}	0.07	0	0	0	0	0
V^{+}	0.09	0	12.8	10.6	23.4	**40.4**
IV^{7}	0.09	0	0	0	0	0
iii^{7}	0.10	5.4	**22.2**	0	0	3.7
iv^{7}	0.16	33.3	7	20.0	22.2	**37.8**
vi^{7}	0.19	3.2	**29.3**	1.0	0	15.6
VI^{7}	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**
Ger^{7}	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
ii^{7}	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad^{6}_{4}	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**
V^{7}	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
Fr^{7}	0.03	0	0	0	0	**57.9**
III^{+}	0.04	0	0	4.1	8.3	**41.7**
i^{7}	0.06	19.0	0	0	**23.2**	0
I^{7}	0.07	0	0	0	0	0
V^{+}	0.09	0	12.8	10.6	23.4	**40.4**
IV^{7}	0.09	0	0	0	0	0
iii^{7}	0.10	5.4	**22.2**	0	0	3.7
iv^{7}	0.16	33.3	7	20.0	22.2	**37.8**
vi^{7}	0.19	3.2	**29.3**	1.0	0	15.6
VI^{7}	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**
Ger^{7}	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
ii^{7}	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad^{6}_{4}	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**
V^{7}	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

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
Fr^7	0.03	0	0	0	0	**57.9**
III^+	0.04	0	0	4.1	8.3	**41.7**
i^7	0.06	19.0	0	0	**23.2**	0
I^7	0.07	0	0	0	0	0
V^+	0.09	0	12.8	10.6	23.4	**40.4**
IV^7	0.09	0	0	0	0	0
iii^7	0.10	5.4	**22.2**	0	0	3.7
iv^7	0.16	33.3	7	20.0	22.2	**37.8**
vi^7	0.19	3.2	**29.3**	1.0	0	15.6
VI^7	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**
Ger^7	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
ii^7	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad^6_4	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**
V^7	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

McLeod and Rohrmeier (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
Fr^7	0.03	0	0	0	0	**57.9**
III^+	0.04	0	0	4.1	8.3	**41.7**
i^7	0.06	19.0	0	0	**23.2**	0
I^7	0.07	0	0	0	0	0
V^+	0.09	0	12.8	10.6	23.4	**40.4**
IV^7	0.09	0	0	0	0	0
iii^7	0.10	5.4	**22.2**	0	0	3.7
iv^7	0.16	33.3	7	20.0	22.2	**37.8**
vi^7	0.19	3.2	**29.3**	1.0	0	15.6
VI^7	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**
Ger^7	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
ii^7	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad^6_4	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**
V^7	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

McLeod and Rohrmeier (2021)
2 out of 31 classes

AugmentedNet
21 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
Fr^{7}	0.03	0	0	0	0	**57.9**
III^{+}	0.04	0	0	4.1	8.3	**41.7**
i^{7}	0.06	19.0	0	0	**23.2**	0
I^{7}	0.07	0	0	0	0	0
V^{+}	0.09	0	12.8	10.6	23.4	**40.4**
IV^{7}	0.09	0	0	0	0	0
iii^{7}	0.10	5.4	**22.2**	0	0	3.7
iv^{7}	0.16	33.3	7	20.0	22.2	**37.8**
vi^{7}	0.19	3.2	**29.3**	1.0	0	15.6
VI^{7}	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**
Ger^{7}	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
ii^{7}	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad^{6}_{4}	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**
V^{7}	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

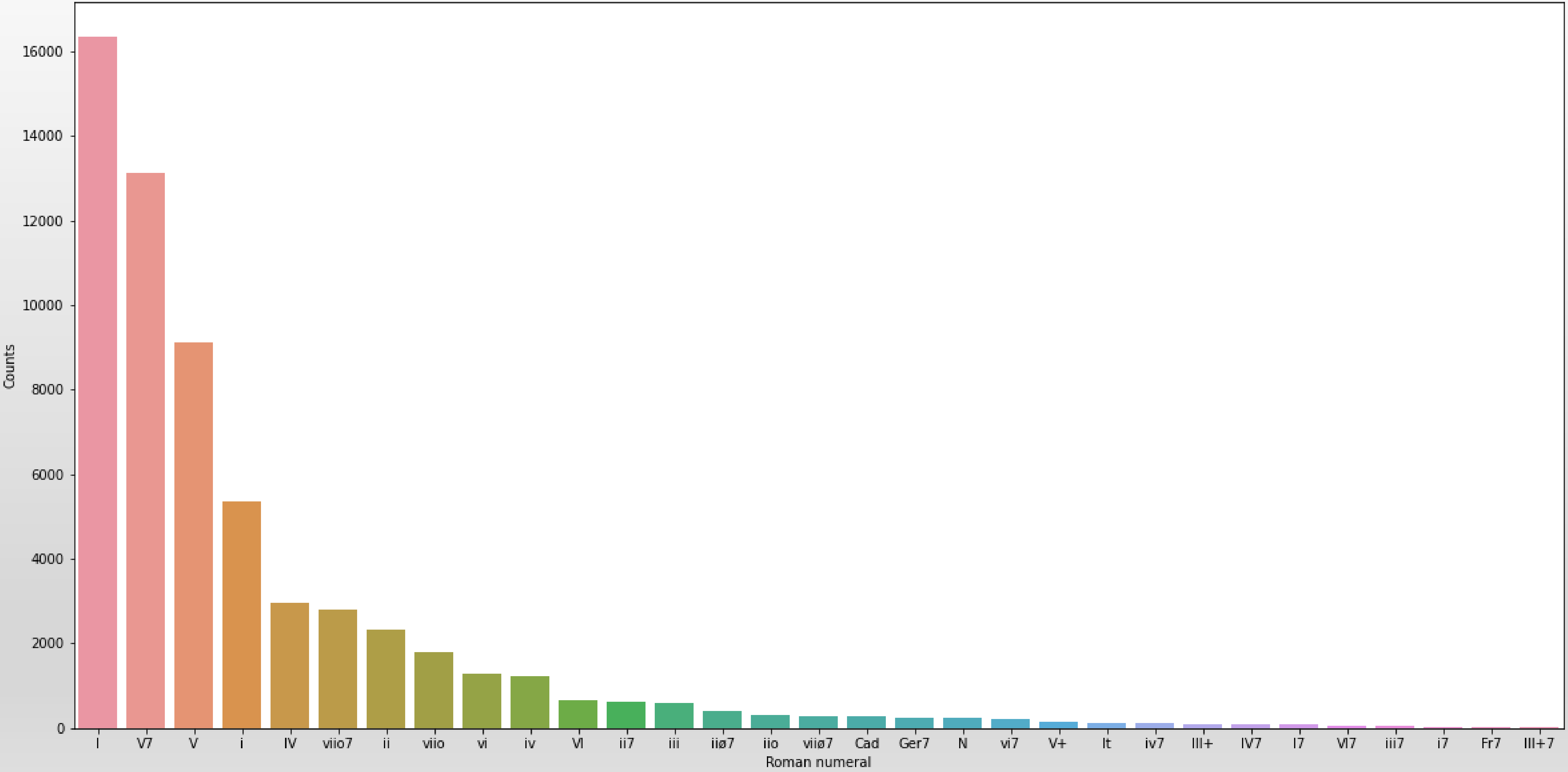
McLeod and Rohrmeier (2021)
2 out of 31 classes

AugmentedNet
21 out of 31 classes

Unrecognized classes
3 out of 31

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
Fr^7	0.03	0	0	0	0	**57.9**
III$^+$	0.04	0	0	4.1	8.3	**41.7**
i^7	0.06	19.0	0	0	**23.2**	0
I^7	0.07	0	0	0	0	0
V$^+$	0.09	0	12.8	10.6	23.4	**40.4**
IV^7	0.09	0	0	0	0	0
iii^7	0.10	5.4	**22.2**	0	0	3.7
iv^7	0.16	33.3	7	20.0	22.2	**37.8**
vi^7	0.19	3.2	**29.3**	1.0	0	15.6
VI^7	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**
Ger^7	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
ii^7	0.91	27.8	37.0	19.1	17.0	**41.0**
Cad$^6_4$	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**
V^7	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**



16000
14000
12000
10000
8000
6000
4000
2000
0
Counts
I
V7
V
i
IV
viio7
ii
viio
vi
iv
VI
ii7
iii
iiø7
iio
viiø7
Cad
Ger7
N
vi7
V+
It
iv7
III+
IV7
I7
VI7
iii7
i7
Fr7
III+7
Roman numeral

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