ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

The Effect of Spectrogram Reconstruction on Automatic Music Transcription: An Alternative Approach to Improve Transcription Accuracy

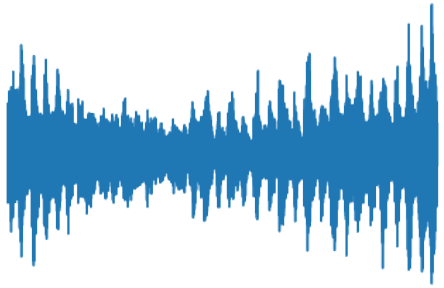
Kin Wai Cheuk[1,2], Yin-Jyun Luo[3], Emmanouil Benetos[3], Dorien Herremans[1,2]

[1] Singapore University of Technology and Design (SUTD)
[2] Agency for Science, Technology and Research (A*STAR)
[3] Queen Mary University of London (QMUL)

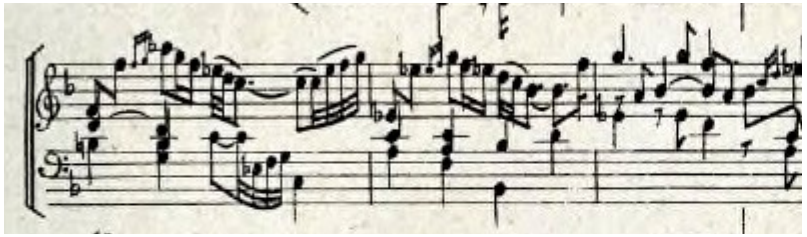
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
Queen Mary University of London
A*STAR Institute of Bioengineering and Nanotechnology

Introduction

Automatic Music transcription (AMT) is the process of transforming audio data into symbolic representations (e.g. music scores)



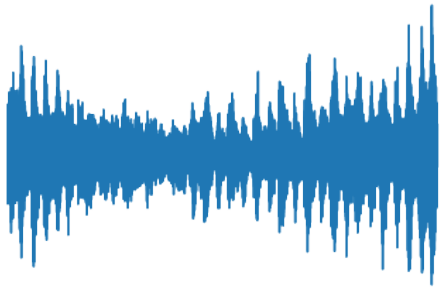
Waveform



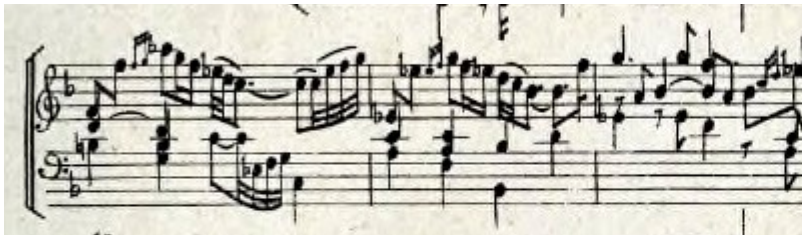
Music score

Introduction

Automatic Music transcription (AMT) is the process of transforming audio data into visual data

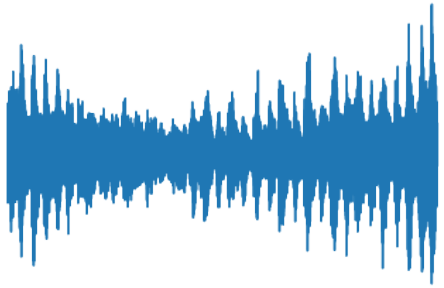


Waveform



Music score

analogous to **Automatic speech recognition (ASR)**



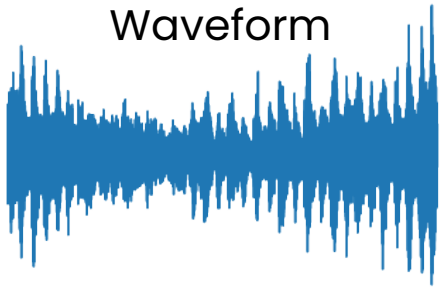
Hello World!

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese
IAPR
UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY
cini
consorzio interuniversitario nazionale per l'informatica

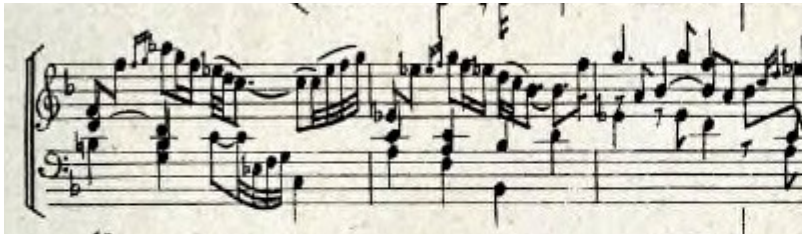
Introduction

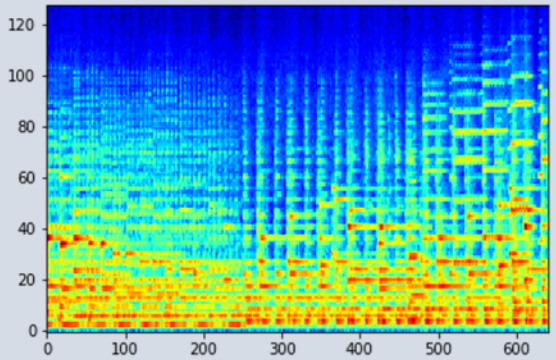
To simplify the problem…

Waveform



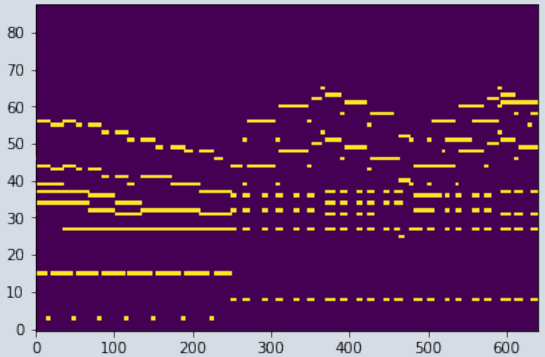
Music score





Mel Spectrogram

≈ Image segmentation



Piano Roll

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese

IAPR

UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media

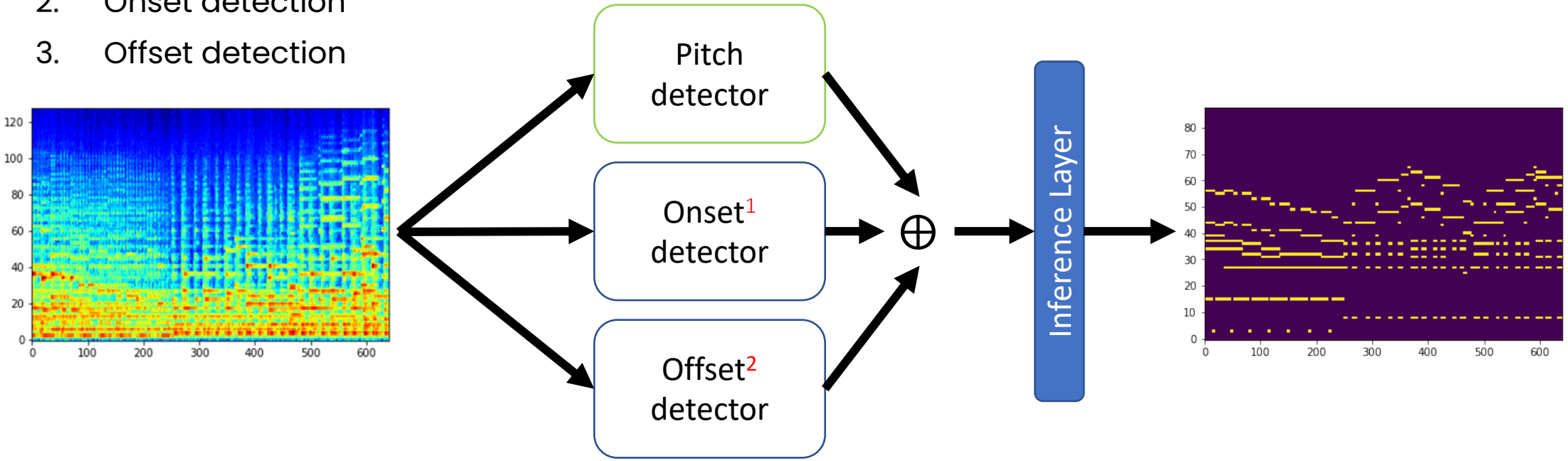
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY

cini
consorzio interuniversitario nazionale per l'informatica

Literature Review

State-of-the-art models usually break down AMT into subtasks:

1. Pitch detection
2. Onset detection
3. Offset detection



[1]Hawthorne et al., “Onsets and Frames: Dual-Objective Piano Transcription”, International Society for Music Information Retrieval Conference 2018 (ISMIR)
[2]Kim and Bello, “Adversarial Learning for Improved Onsets and Frames Music Transcription”, ISMIR 2019

ICPR 2020 25th International Conference on Pattern Recognition Milan, Italy 10 | 15 January 2021
CVPL Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)
UNIMORE Università degli Studi di Modena e Reggio Emilia
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese
IAPR
Università degli Studi Firenze
MICC Centro per la Comunicazione e l'Integrazione dei Media
Technically Co-Sponsored by IEEE
IEEE Computer Society
cini consorzio interuniversitario nazionale per l'informatica

Model Description

nnAudio: toolbox for GPU spectrogram extraction

We use U-net as the foundation of our model, since U-net performs well on image segmentation.

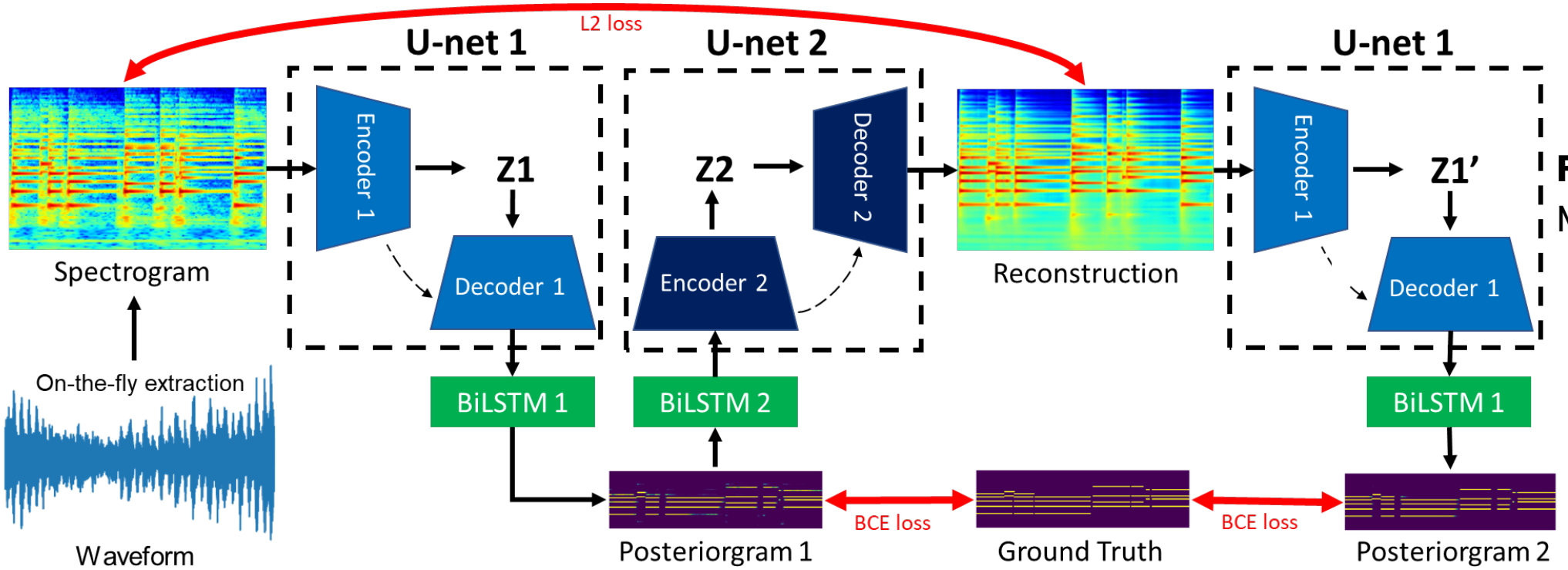


Fig. 1 of the paper
Model architecture

Feedforward
Skip connection
Loss function

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese
IAPR
UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY
cini
consorzio interuniversitario nazionale per l'informatica

Model Description

nnAudio: toolbox for GPU spectrogram extraction

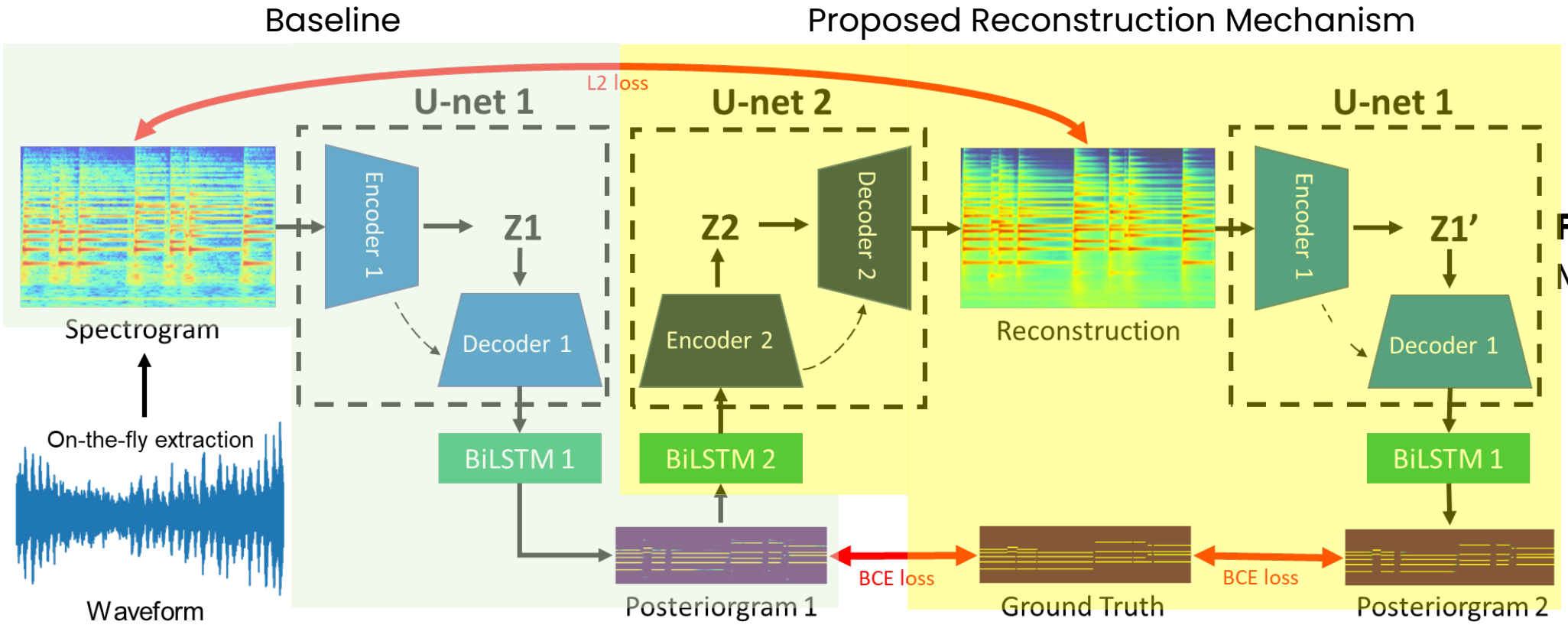


Fig. 1 of the paper
Model architecture

Feedforward
Skip connection
Loss function

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese
IAPR
UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY
cini
consorzio interuniversitario nazionale per l'informatica

Results: MAPS dataset

	Frame Level			Note Level			Note with offset		
	P	R	F1	P	R	F1	P	R	F1
Kelz et at.	81.2	65.1	71.6	44.3	61.3	50.9	20.1	27.8	23.1
Hawthorne (Frame only)	-	-	**76.1**	-	-	62.7	-	-	27.9
Hawthorne 2018	88.5	70.9	78.3	84.2	80.7	82.3	51.3	49.3	50.2
CQT (baseline)	79.7±7.0	**67.7**±9.0	72.9±7.3	57.9±11.1	57.2±11.9	57.2±11.0	34.7±11.1	34.4±11.8	34.4±11.3
Mel (baseline)	84.7±6.0	67.2±9.7	74.5±7.2	60.2±11.3	60.5±12.2	60.1±11.2	36.3±10.7	**36.7**±11.8	36.3±11.0
CQT (proposed)	86.3±5.9	61.4±11.8	71.2±9.3	67.8±10.9	57.7±12.9	41.9±11.4	39.7±10.9	34.0±11.5	34.4±11.0
Mel (proposed)	**89.3**±5.5	61.9±11.1	72.5±8.7	**71.3**±9.5	**62.7**±12.5	**66.3**±10.6	**41.3**±11.3	36.5±12.2	**38.5**±11.6

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese
IAPR
UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY
cini
consorzio interuniversitario nazionale per l'informatica

Results: MAESTRO dataset

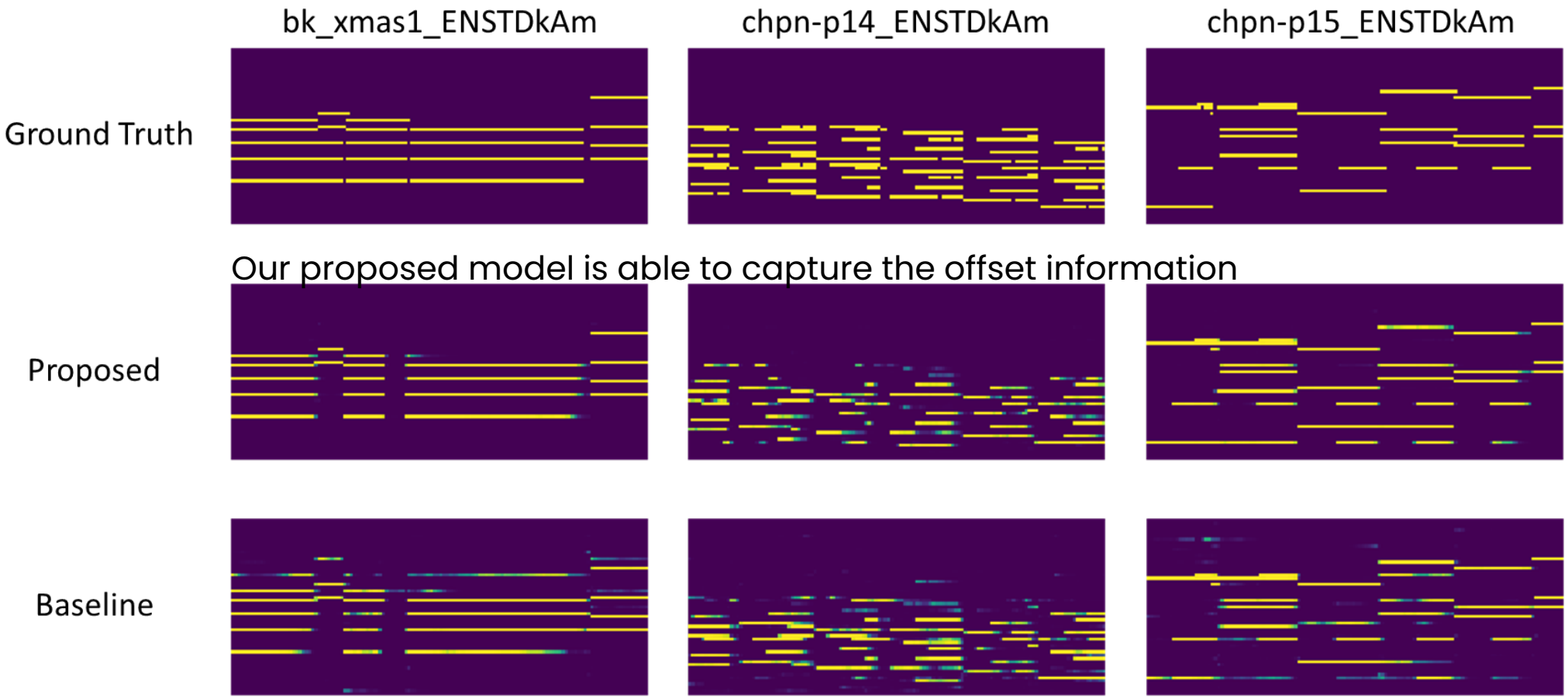
	Frame Level			Note Level			Note with offset		
	P	R	F1	P	R	F1	P	R	F1
Hawthorne 2019	92.9	78.5	84.9	87.5	85.6	86.4	66.2	66.8	67.4
Kim 2019	93.1	89.8	91.4	98.1	93.2	95.6	94.1	78.1	81.0
CQT (baseline)	91.3±3.3	65.8±10.0	76.1±7.3	68.8±11.4	65.9±11.9	67.0±10.9	36.7±10.5	35.2±10.5	35.8±10.2
Mel (baseline)	90.2±3.5	**71.4**±10.3	**79.4**±7.1	68.4±11.6	65.1±13.2	66.5±11.9	42.0±10.7	**40.1**±11.5	40.9±10.9
CQT (proposed)	89.1±4.3	67.3±10.8	76.1±7.6	72.6±11.8	63.7±12.9	67.5±11.7	44.5±10.9	39.5±12.0	41.6±11.3
Mel (proposed)	**94.0**±2.7	66.1±12.4	77.0±9.0	**78.9**±9.4	**68.9**±12.6	**73.3**±10.7	**44.8**±10.6	39.3±11.6	**41.7**±11.0

Results: MusicNet dataset

	Frame Level			Note Level			Note with offset		
	μAP	A	F1	P	R	F1	P	R	F1
Thickstun 2019	**77.3**	**55.3**	-	-	-	-	-	-	-
Pedersoli 2020	75.6	-	-	-	-	-	-	-	-
CQT (baseline)	69.6±9.4	51.1±8.3	67.2±7.6	60.1±11.7	50.3±21.4	53.8±17.4	32.7±13.4	28.6±18.1	30.0±16.2
Mel (baseline)	71.1±12.3	53.1±11.2	**68.6**±10.1	59.6±12.3	49.3±22.3	53.0±18.4	31.2±12.1	26.9±16.6	28.4±14.8
CQT (proposed)	71.1±14.8	42.9±15.1	58.4±15.5	**64.2**±13.6	**51.8**±25.9	55.2±21.9	**37.7**±17.3	30.4±21.8	32.5±20.0
Mel (proposed)	71.1±14.6	50.7±12.6	66.3±11.9	63.5±14.0	51.4±23.9	**55.7**±20.2	37.7±18.8	**32.2**±23.2	**34.1**±21.6

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese
IAPR
UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY
cini
consorzio interuniversitario nazionale per l'informatica

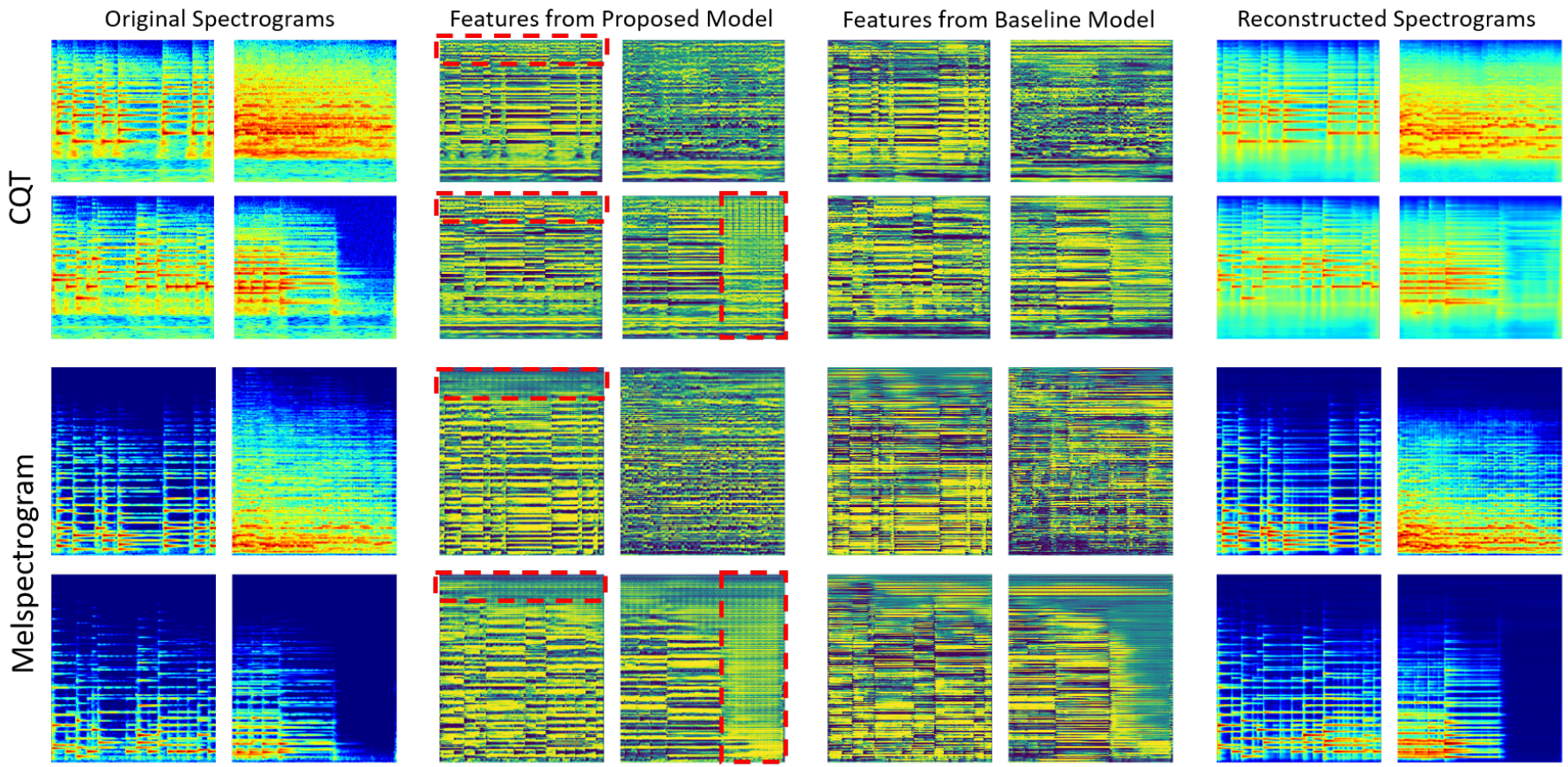
Transcription Results



Our proposed model is able to capture the offset information

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese
IAPR
UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media
Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY
cini
consorzio interuniversitario nazionale per l'informatica

Learned features



ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese

IAPR

UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media

Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY

cini
consorzio interuniversitario nazionale per l'informatica

Summary

- The reconstructed spectrogram is a denoised version of the original spectrogram
- Transcribing from the denoised spectrogram results in a better posteriorgram
- The posteriorgram shows features that our proposed model is aware of the offset locations
- The feature learned by our model shows grid-like structures. LSTM might be able to use this feature to figure out the onset and offset locations

Future work

- We will extend our model to predict also onset and offset, and use these predictions together with the frame prediction to predict the final posteriorgram
- Our proposed model might serve as a framework for semi-supervised learning
- The LSTM layers can be replaced with self-attention layers

Acknowledgment

This work is supported by:

- Agency for Science, Technology and Research (A*STAR) under grant no. SING-2018-02-0204
- MOE Tier 2 grant no. MOE2018-T2-2-161
- SRG ISTD 2017 129

Source code

nnAudio: toolbox for GPU spectrogram extraction

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

CVPL
Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning (CVPL- ex-GIRPR)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Centro Interdipartimentale di Ricerca Industriale SOFTECH: ICT per le Imprese

IAPR

UNIVERSITÀ DEGLI STUDI FIRENZE
MICC
Centro per la Comunicazione e l'Integrazione dei Media

Technically Co-Sponsored by
IEEE
IEEE COMPUTER SOCIETY

cini
consorzio interuniversitario nazionale per l'informatica