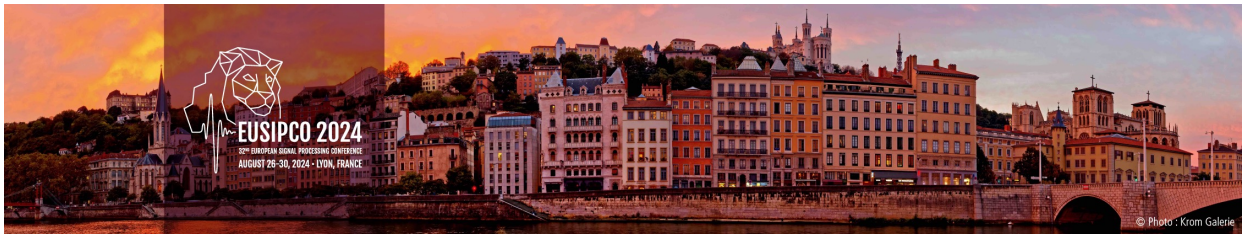
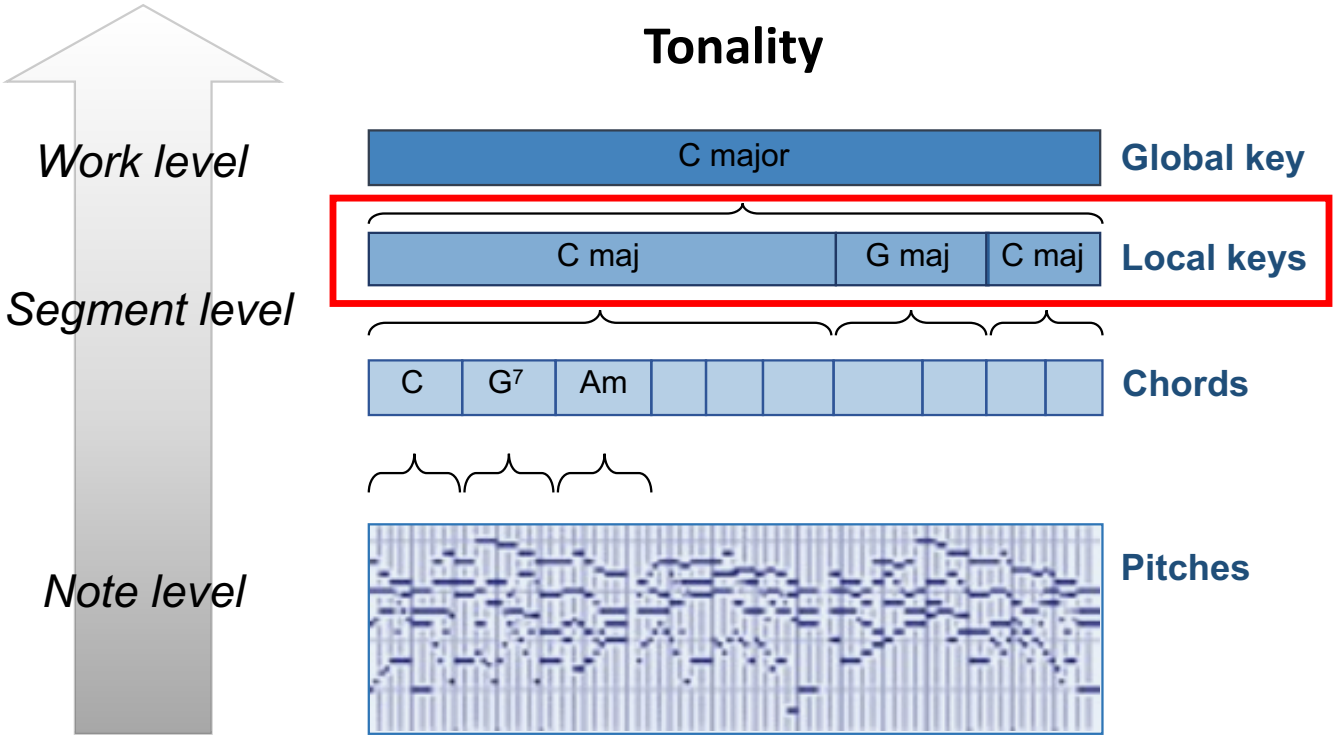


Towards Robust Local Key Estimation with a Musically Inspired Neural Network

Yiwei Ding and Christof Weiß
Computational Humanities
Center for Artificial Intelligence and Data Science, Universität Würzburg
yiwei.ding@uni-wuerzburg.de

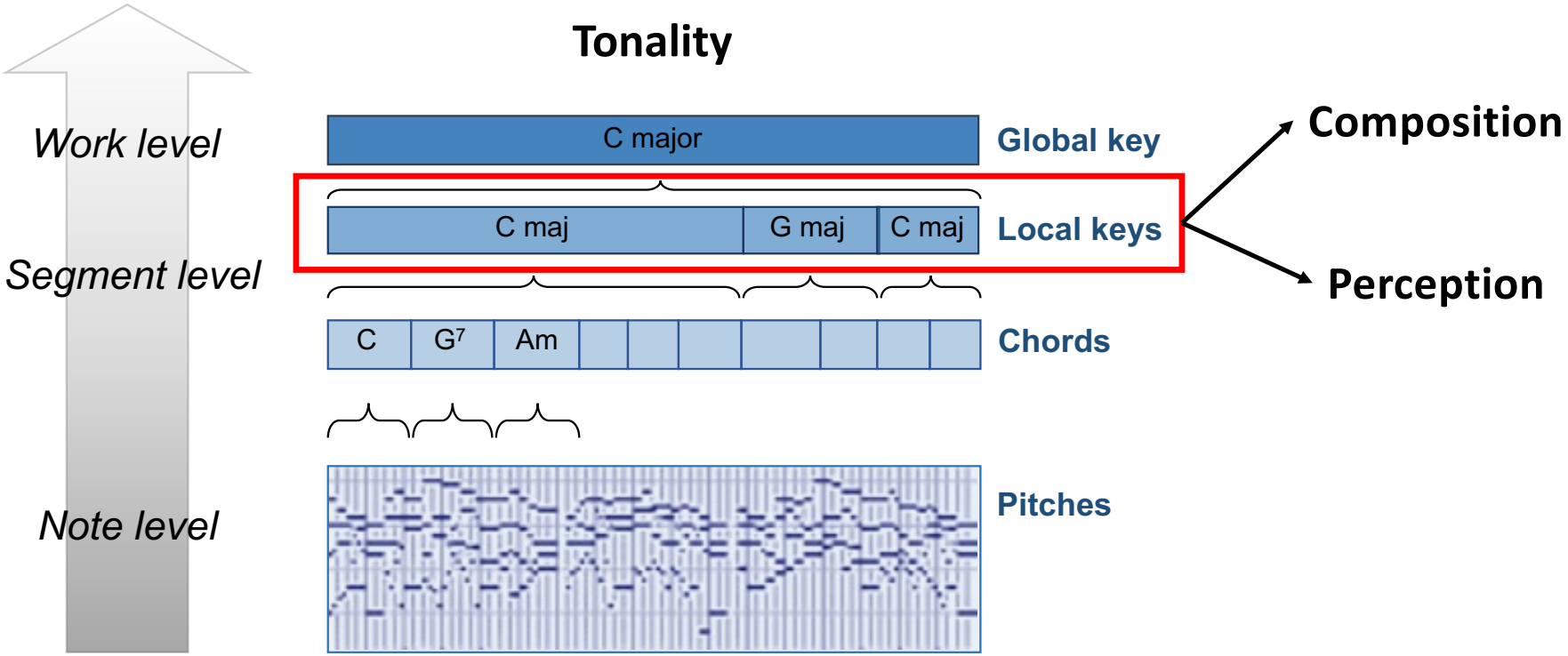
Introduction

Local key



Introduction

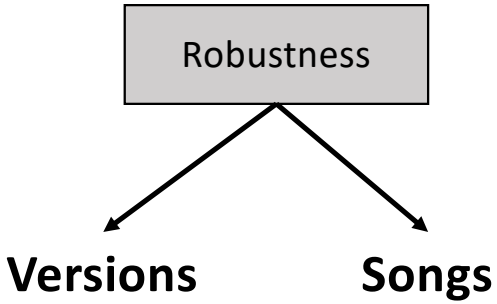
Local key

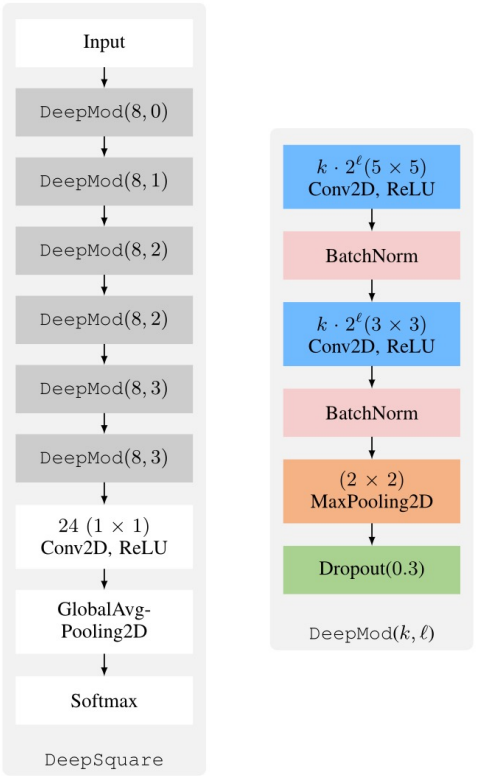


Related Work

Local Key Estimation

- HMM-based
- CNN-based



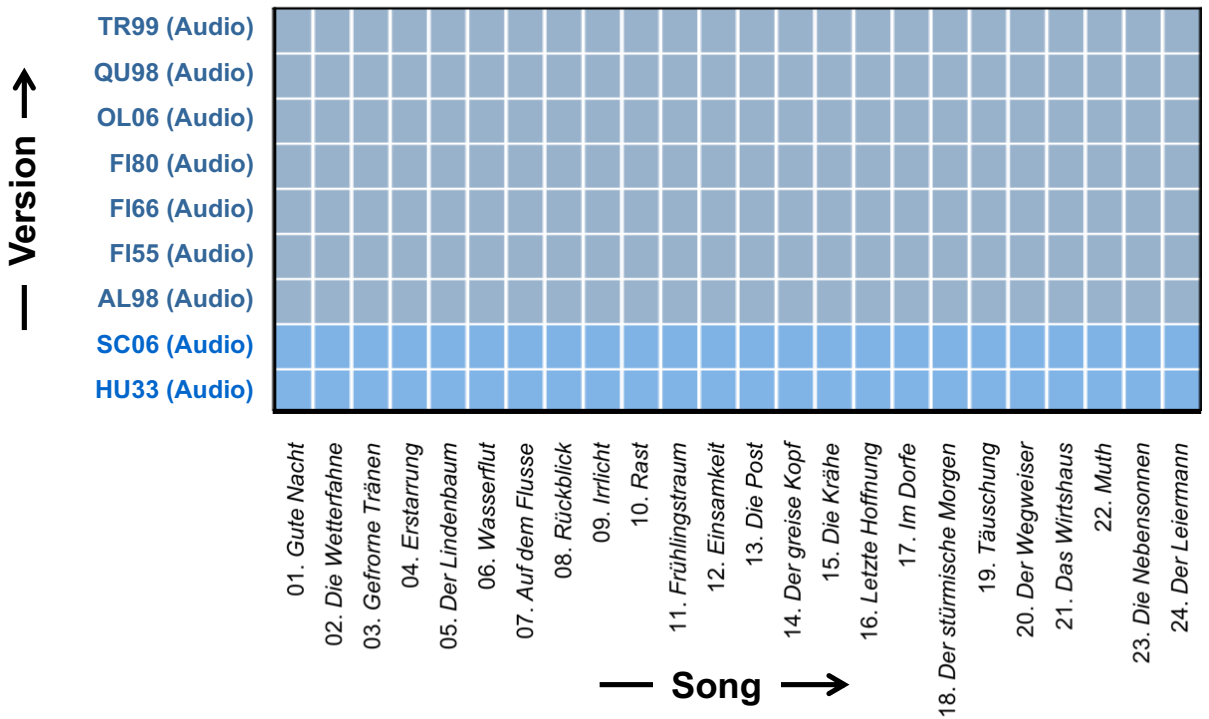


Weiß et. al., ”Local Key Estimation in Music Recordings: A Case Study Across Songs, Versions and Annotators”, *IEEE/ACM TASLP*, 2020.

Dataset

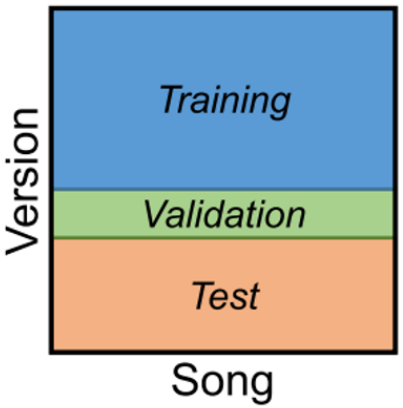
Schubert Winterreise

- Cross-version dataset
- 24 songs * 9 versions

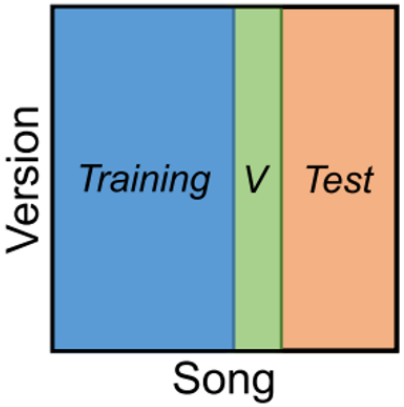


Dataset

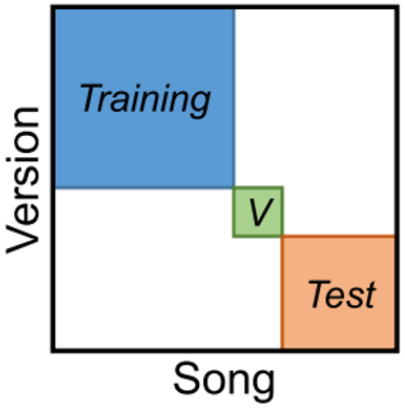
Data splits



Version Split
(easy)



Song Split



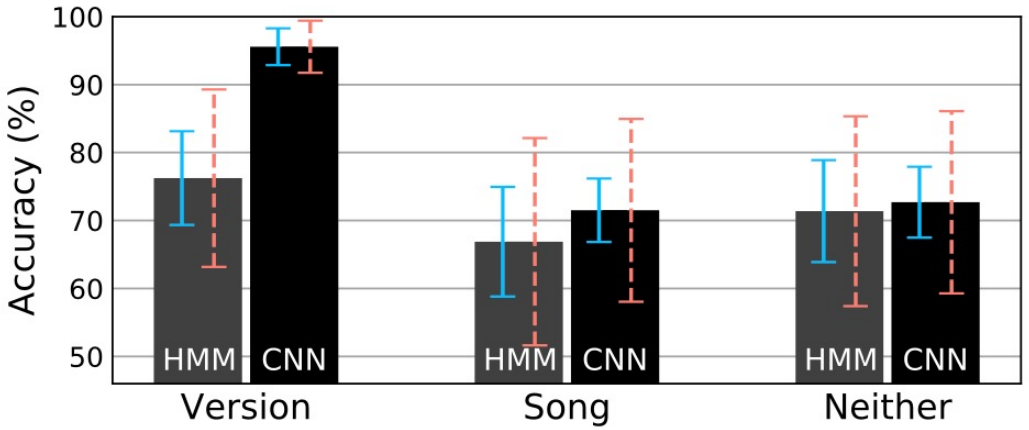
Neither Split
(realistic)

~59% data unused!

Related Work

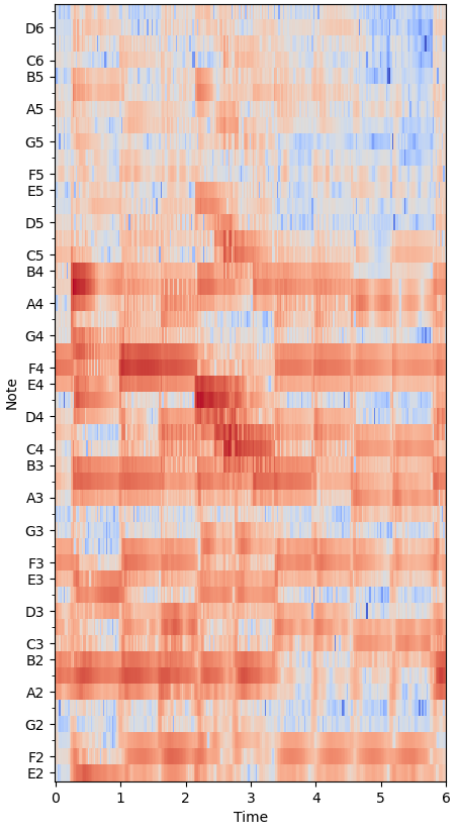
Local Key Estimation

- CNN outperforms HMM on all splits
- Generalization to unseen versions is easier



Proposed Method

OctaveNet



A4

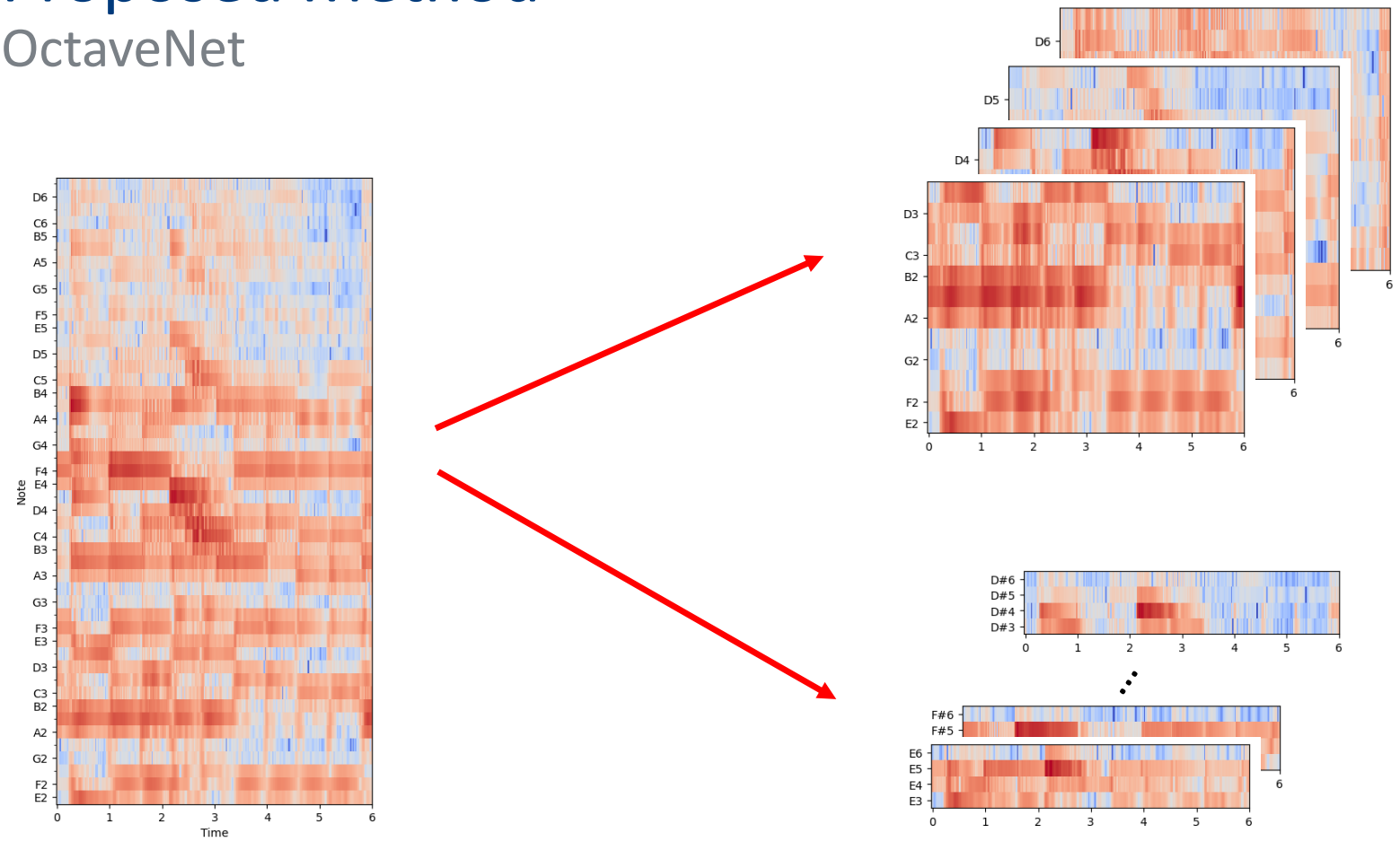
A3

Far away!

CQT Representation

Proposed Method

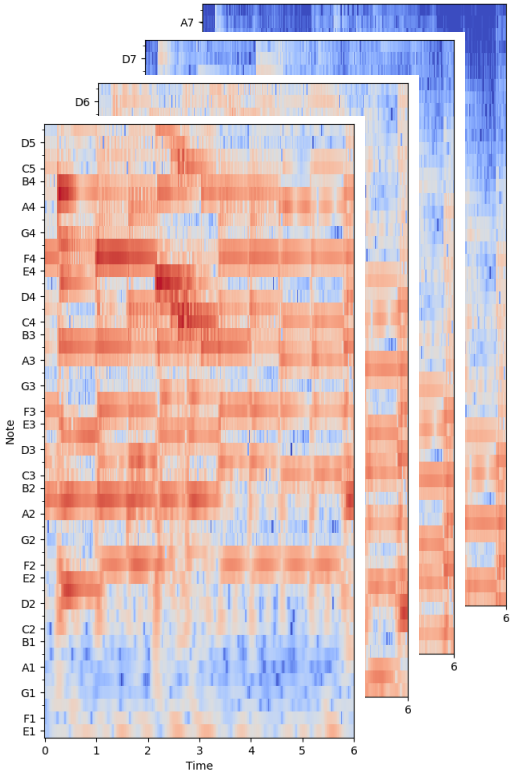
OctaveNet



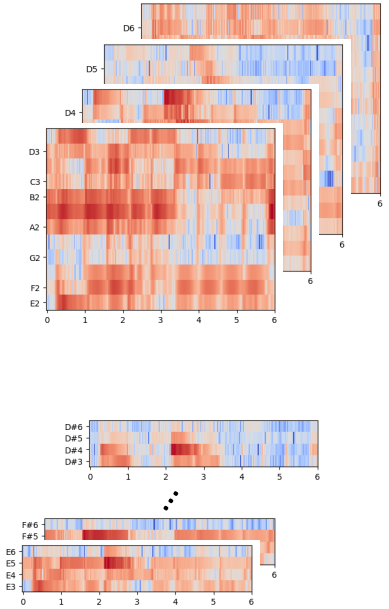
CQT Rearrangement

Proposed Method

OctaveNet vs HCQT



HCQT



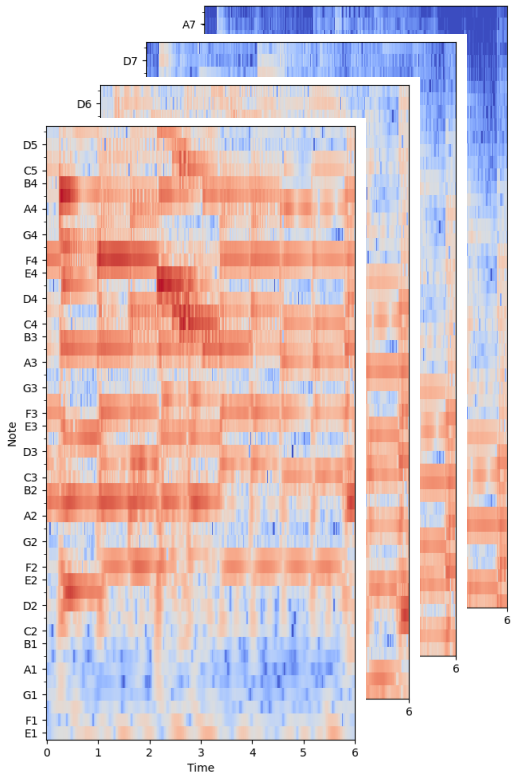
OctaveNet

Bittner et. al., ”Deep Salience Representations for F0 Estimation In Polyphonic Music”, *in Proc. of ISMIR*, 2017.

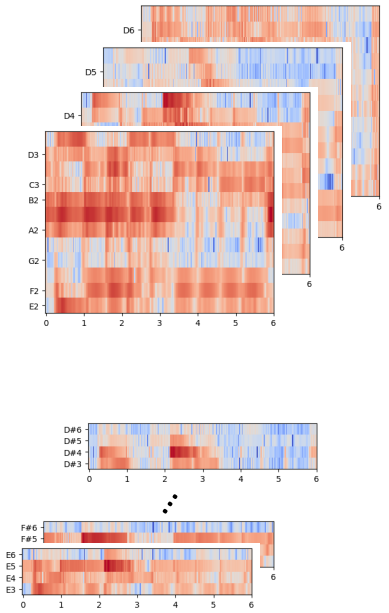
Proposed Method

OctaveNet vs HCQT

- Harmonics-level /Octave-level
- Multiple CQT / Single CQT



HCQT

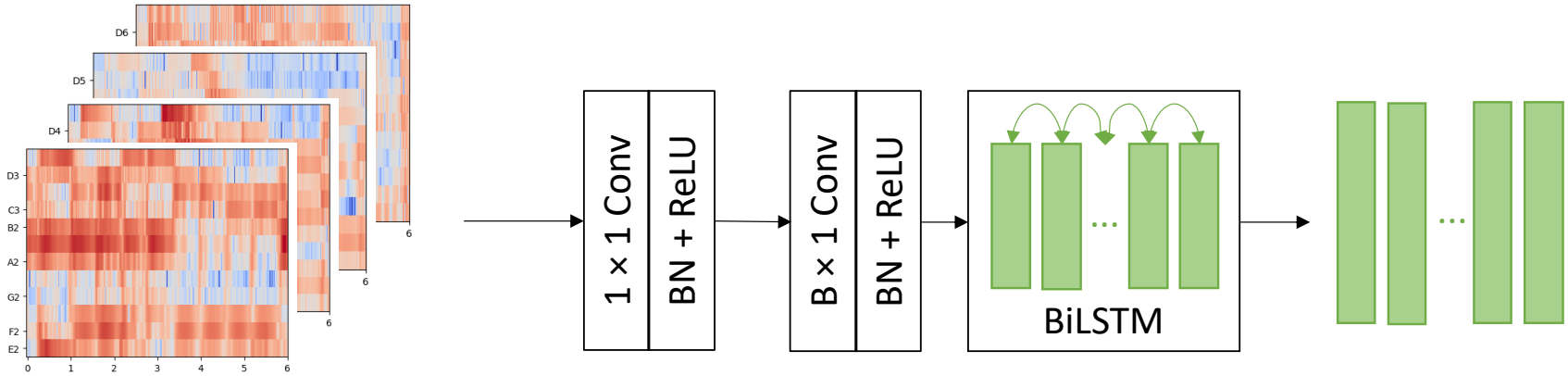


OctaveNet

Bittner et. al., ”Deep Salience Representations for F0 Estimation In Polyphonic Music”, *in Proc. of ISMIR*, 2017.

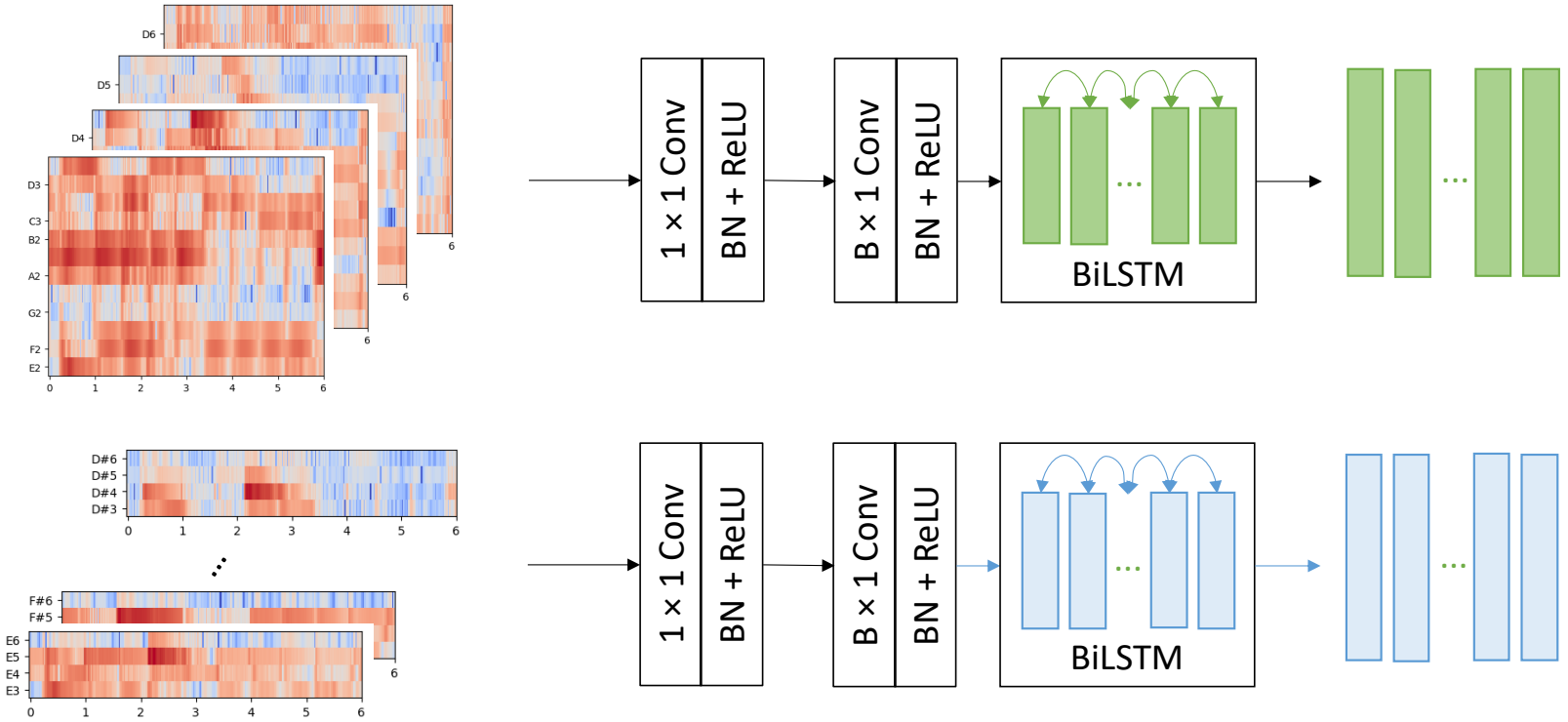
Proposed Method

OctaveNet



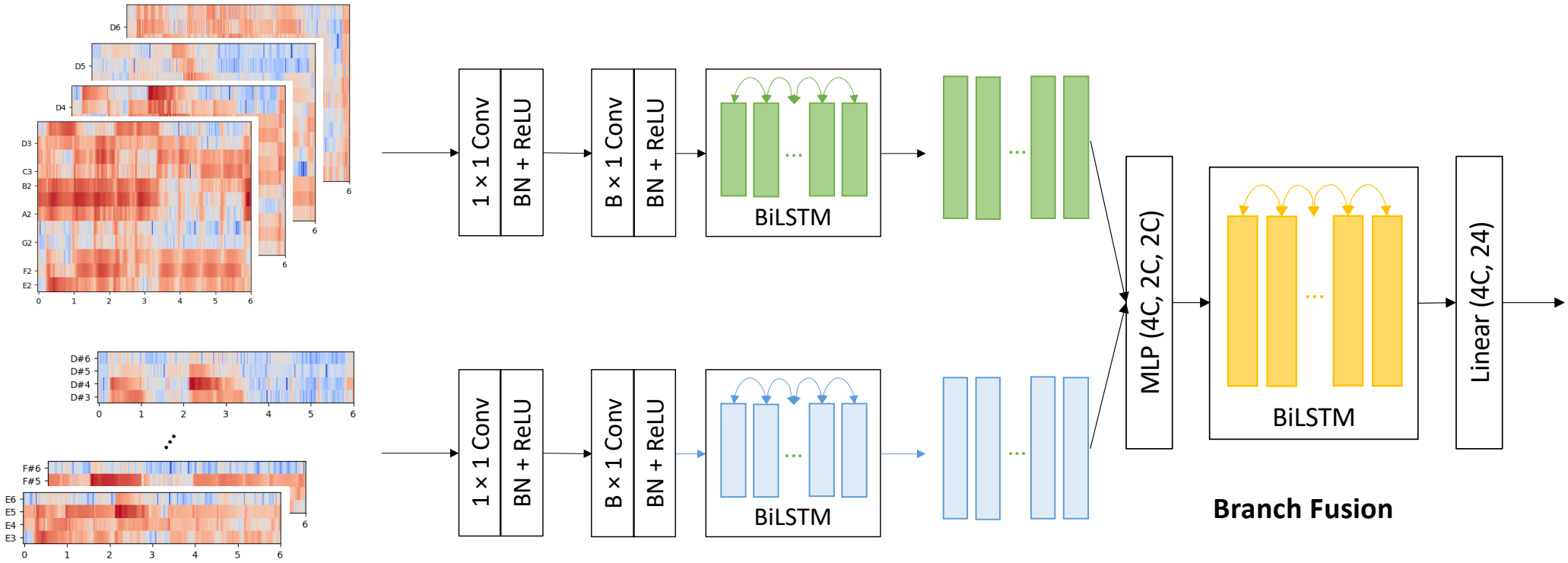
Proposed Method

OctaveNet



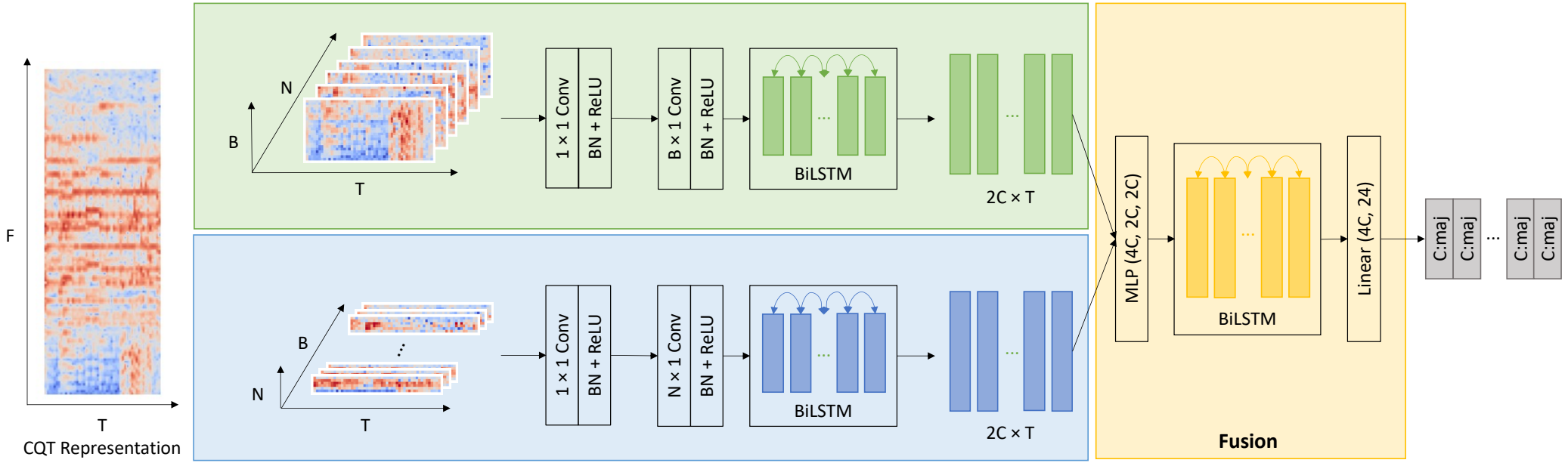
Proposed Method

OctaveNet



Proposed Method

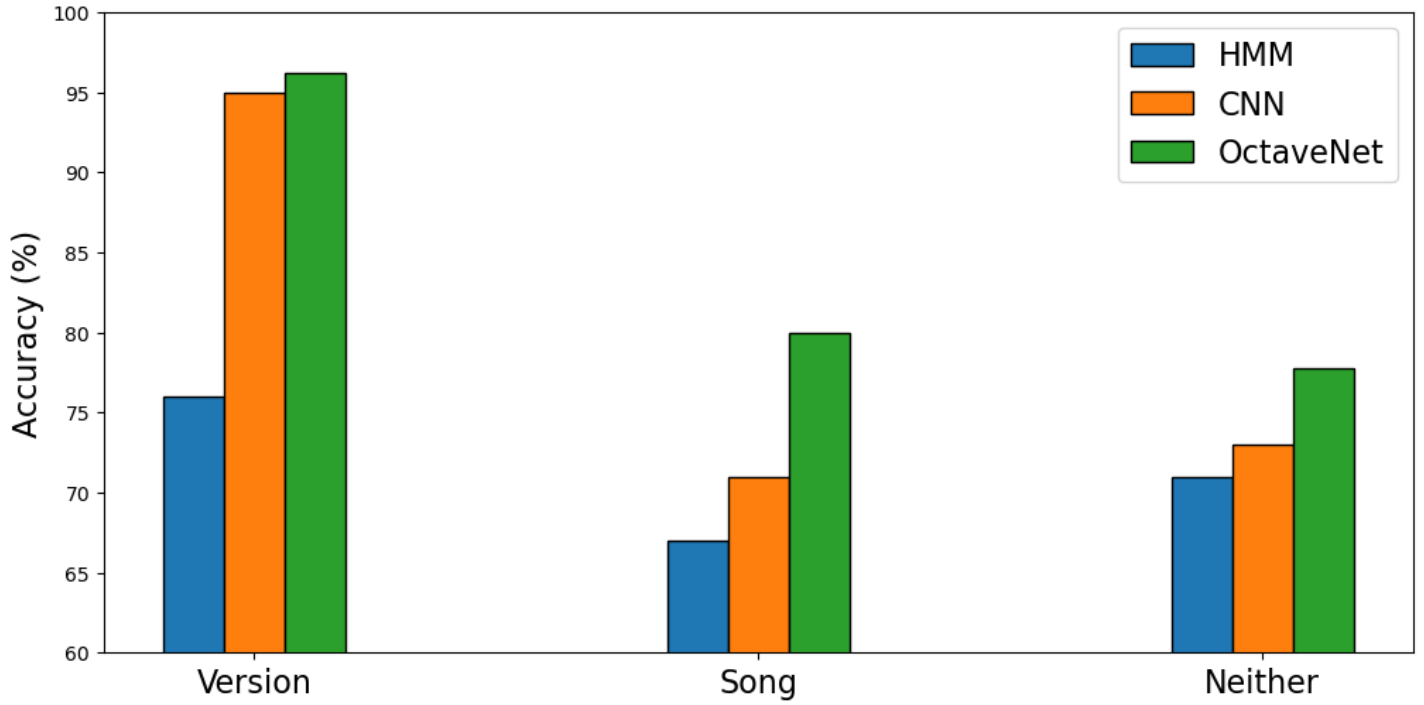
OctaveNet



Results

Comparison with baselines

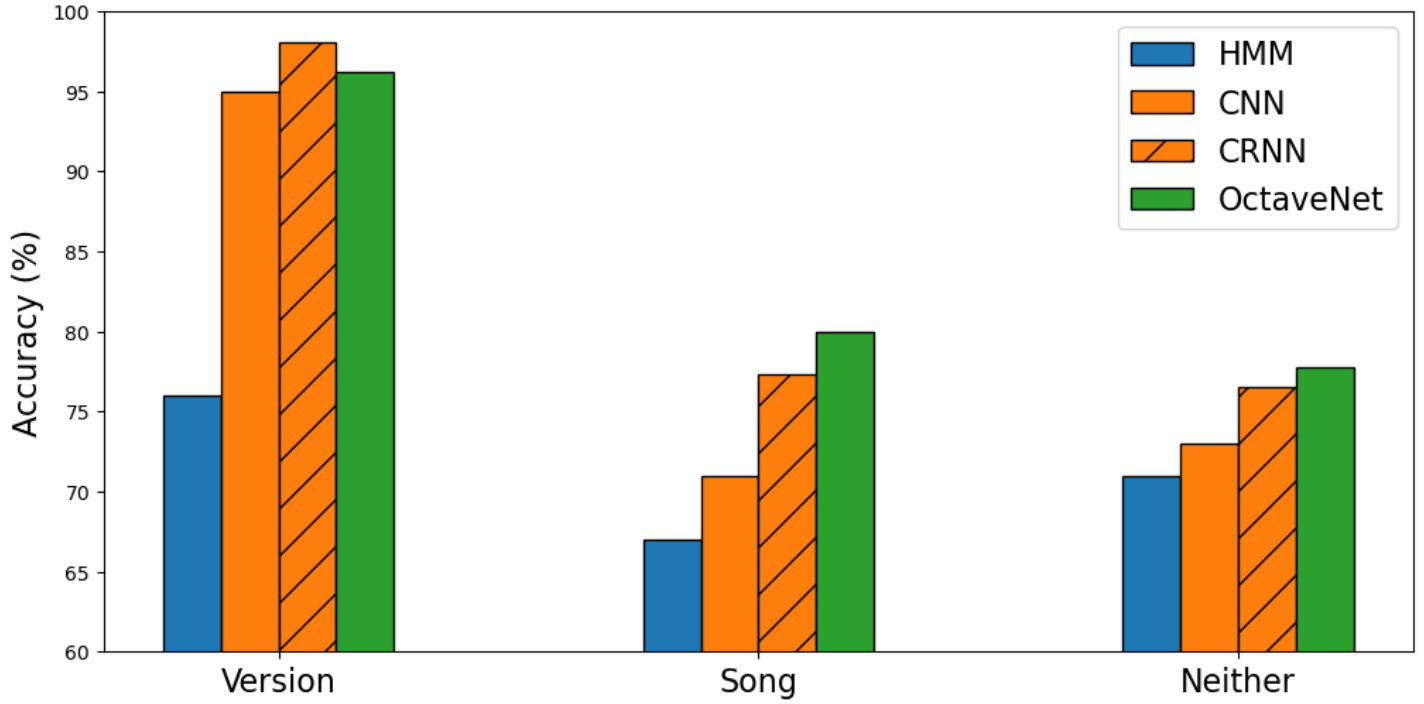
- OctaveNet outperforms the baselines on all splits



Results

Ablation studies

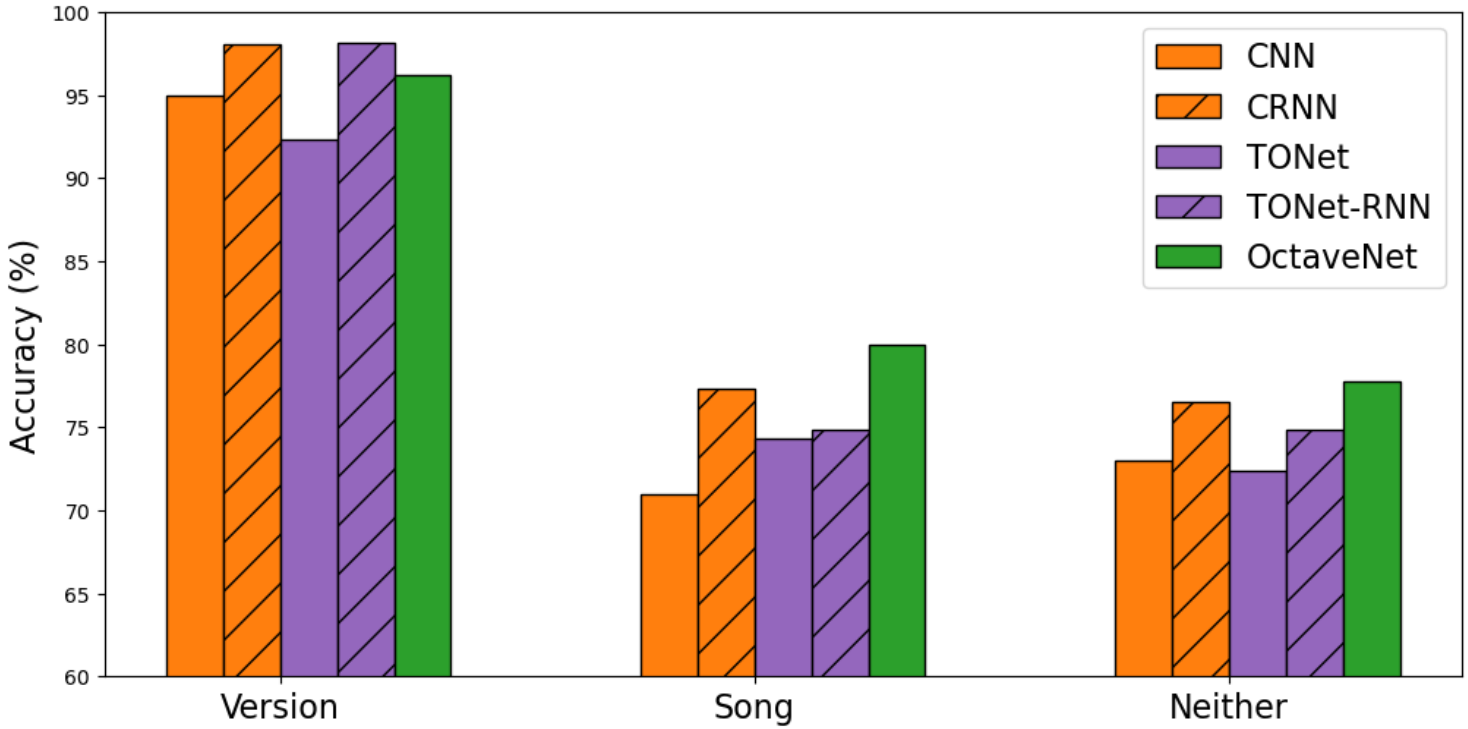
- CRNN performs better than CNN -> Recurrent structure is useful
- OctaveNet outperforms CRNN -> CQT rearrangement is useful



Results

Ablation studies

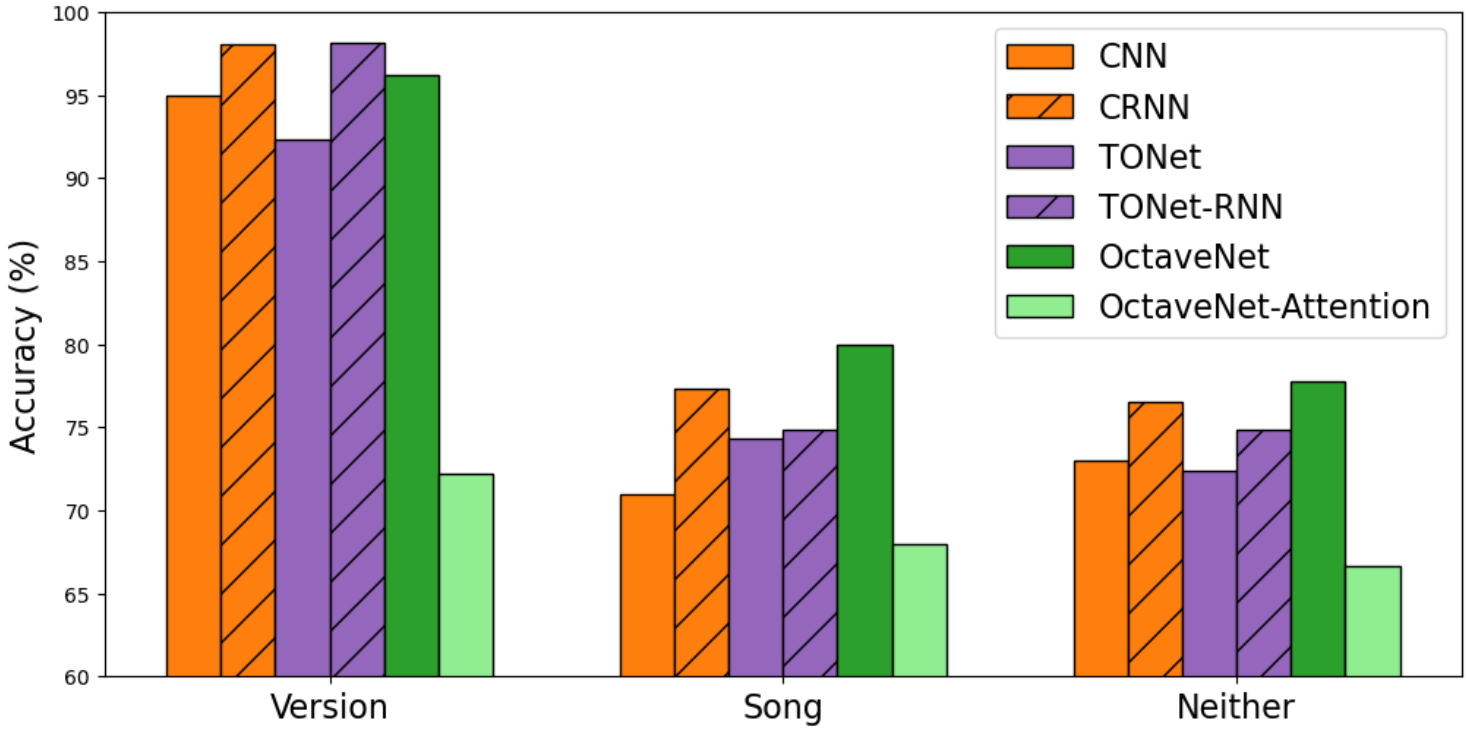
- OctaveNet is also better than TONet (w/ or w/o RNN) on song/neither split



Results

Ablation studies

- Self-attention does not work as well as RNN



Results

Comparison with baselines & ablation studies

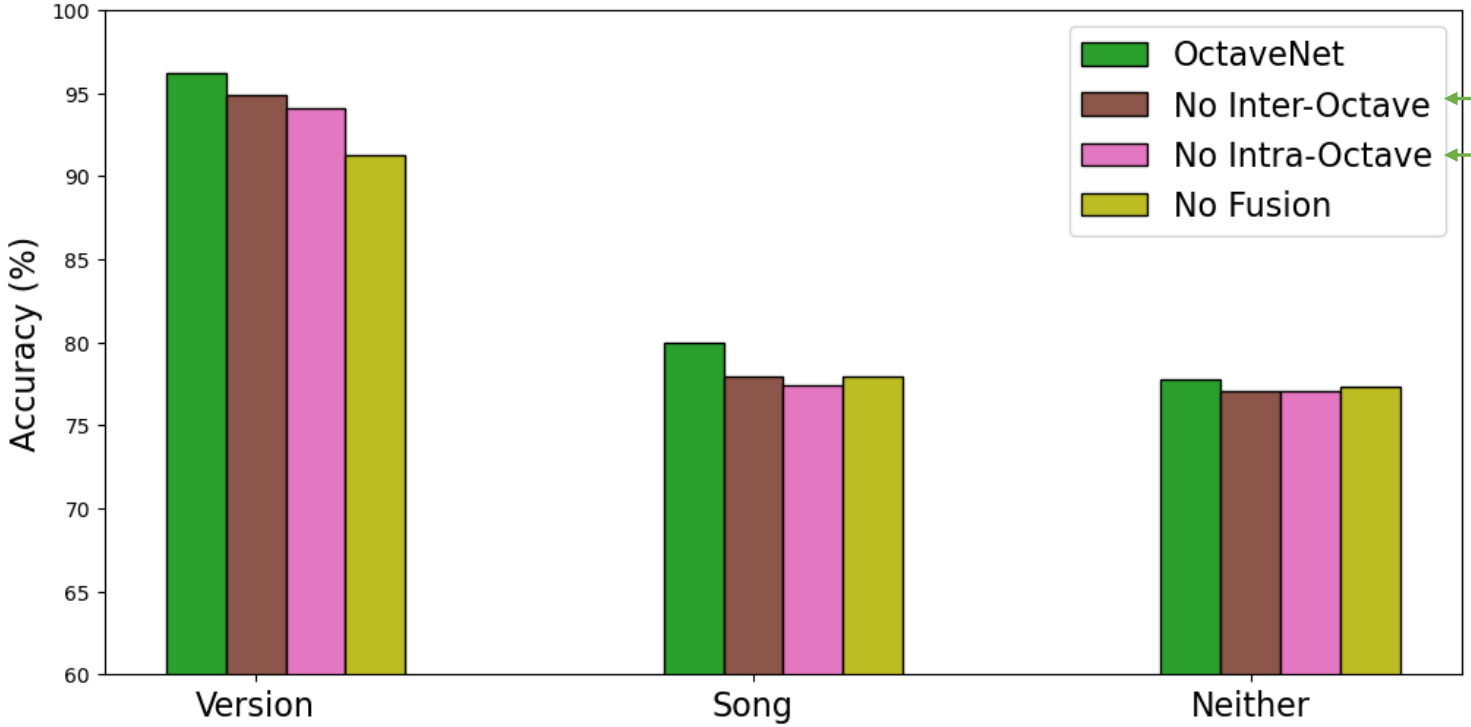
- Less parameters
- Similar or better results

Model	# Params.	Version	Song	Neither
HMM [5]	—	76*	67*	71
CNN [5]	293,296	95*	71*	73
CRNN	361,392	98.03	77.32	76.49
TONet [18]	611,528	92.34	74.28	72.37
TONet-RNN	722,760	**98.17**	74.87	74.85
OctaveNet	149,720	96.23	**80.00**	**77.75**

Results

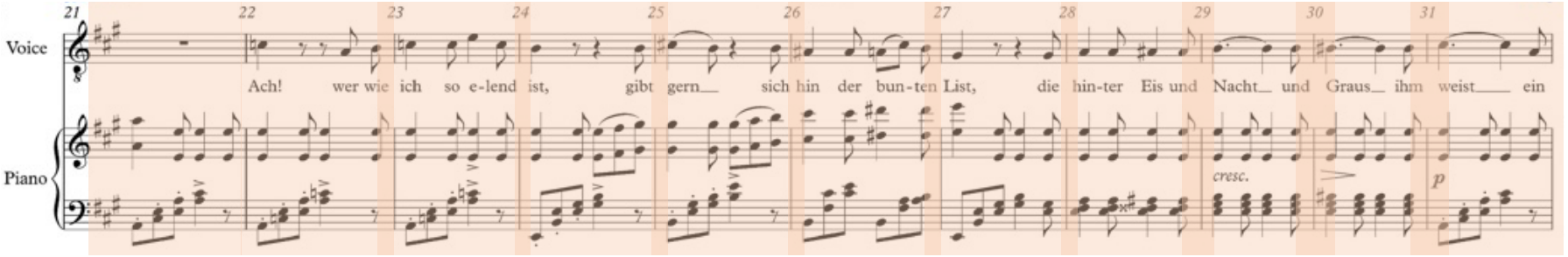
Ablation studies

- Removing any component leads to a performance decay
- …Really? -> Performance are similar on the neither split!



Results

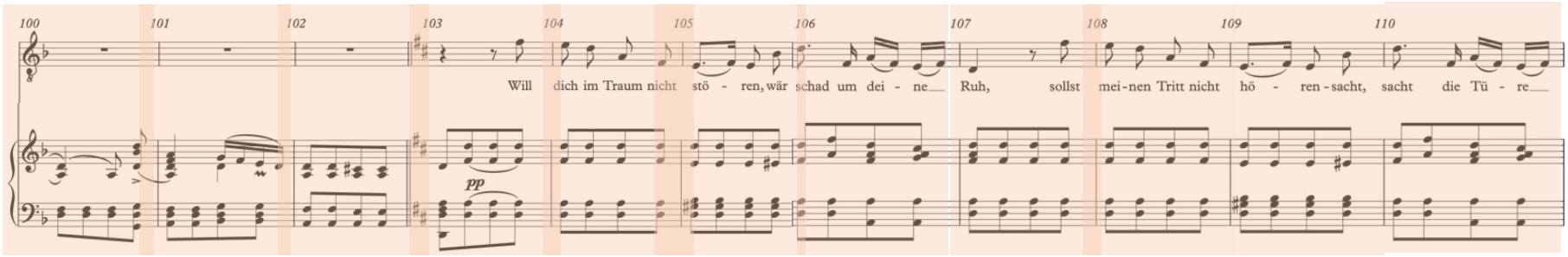
Case study: success

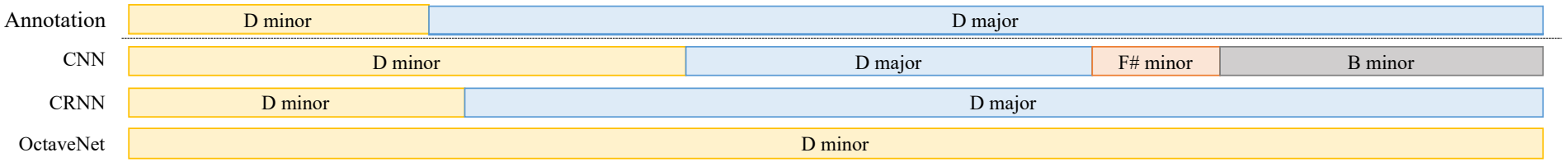


Annotation	A major
CNN	A major | E major
CRNN	A major | A minor | E major | A major
OctaveNet	A major

Results

Case study: failure





Conclusion

- We propose **OctaveNet**, a musically inspired neural network for local key estimation.
- With **less parameters**, it outperforms several baselines on the Schubert Winterreise dataset.
- The domain knowledge in the **CQT rearrangement** and the **recurrent components** are two keys for the improvement.

Code (GitHub repo)

Computational Humanities

Personal Website😜