

Utilizing Cross-Version Consistency for Domain Adaptation: A Case Study on Music Audio

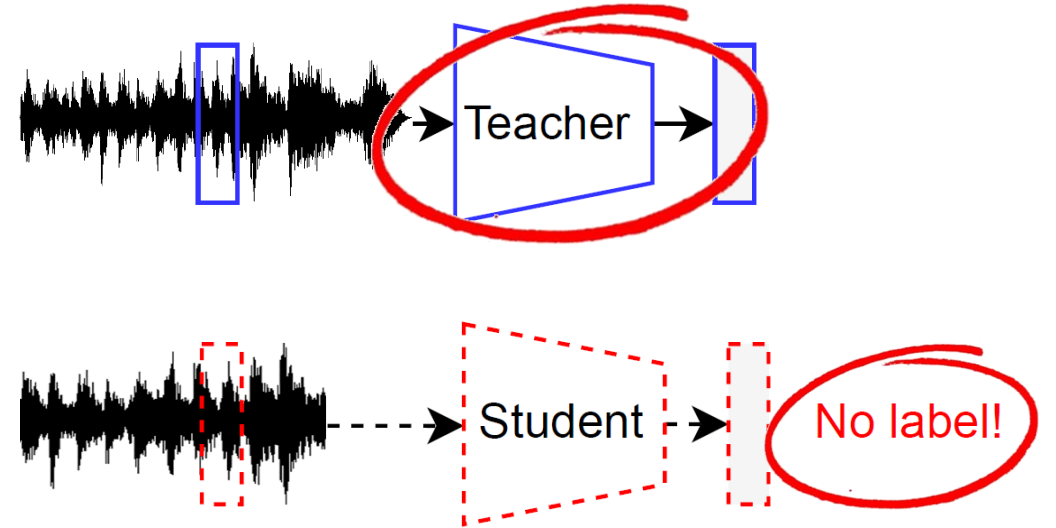
Lele Liu and Christof Weiß

Center for Artificial Intelligence and Data Science, University of Würzburg
{lele.liu, christof.weiss}@uni-wuerzburg

Unsupervised Domain Adaptation

Teacher–Student Approach

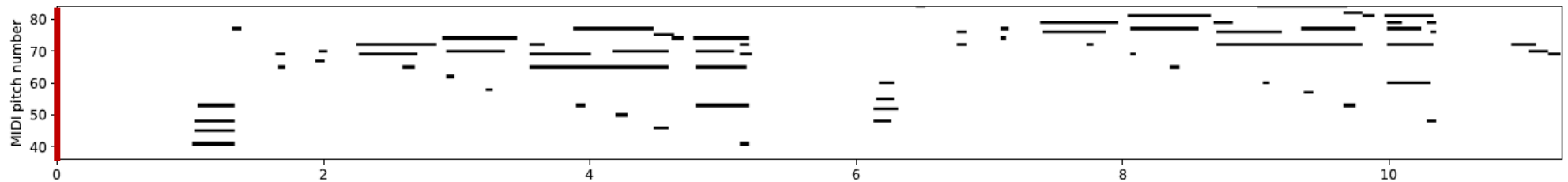
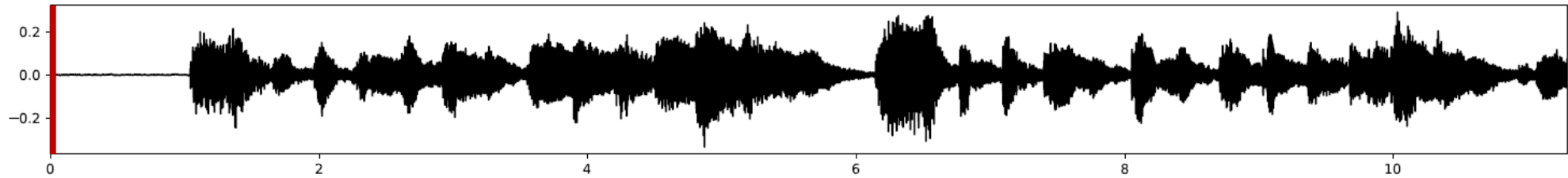
- **Teacher** model (**source domain**), supervised learning
- Want: **Student** model (**target domain**, unseen by teacher)
- Teacher–student approach:
Use teacher to **generate pseudo-labels**



Teacher Domain

Piano Solo (MAESTRO dataset)

- Johann Sebastian Bach, *Italian Concerto* BWV 971 in F Major



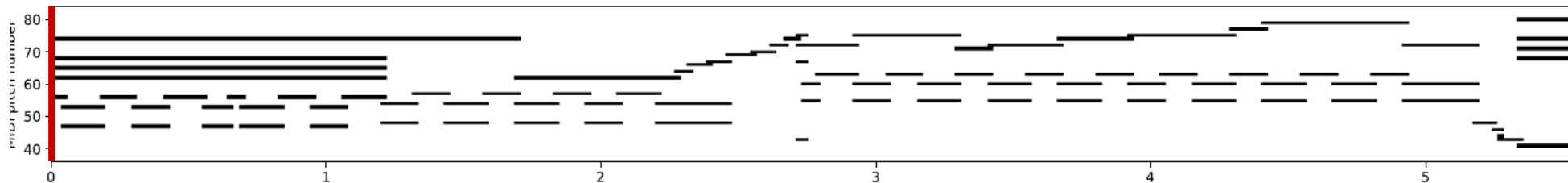
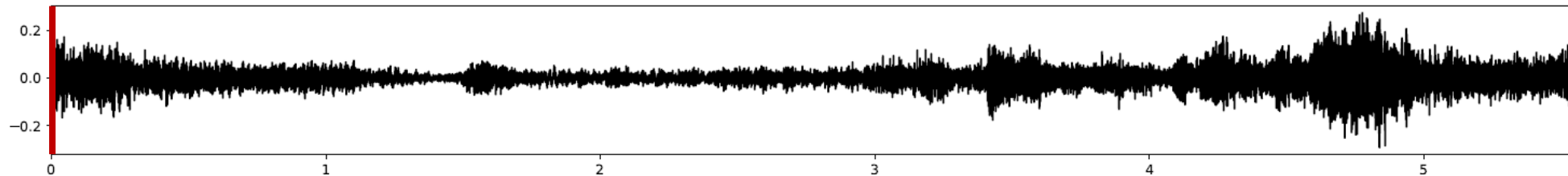
Allegro animato

The musical score is written in 2/4 time with a key signature of one flat (F major). It features a treble and bass clef. The first measure starts with a forte (*f*) dynamic. The score includes various musical notations such as eighth notes, quarter notes, and chords. Fingerings are indicated by numbers 1-5 above the notes. The second measure has a forte (*f*) dynamic, and the third measure has a mezzo-forte (*mf*) dynamic. The score ends with a mezzo-forte (*mf*) dynamic.

Student Domains

Singing + Orchestra (Wagner Ring Dataset)

- Richard Wagner: *Der Ring des Nibelungen, Die Walküre* WWV 86 B3, Act 3



ORTLINDE

WALTRAUTE Den Fels er-reich-ten Roß und Rei-ter!

WOTAN Den Fels er-reich-ten Roß und Rei-ter!

Steh! Brunn-hild! Hbl

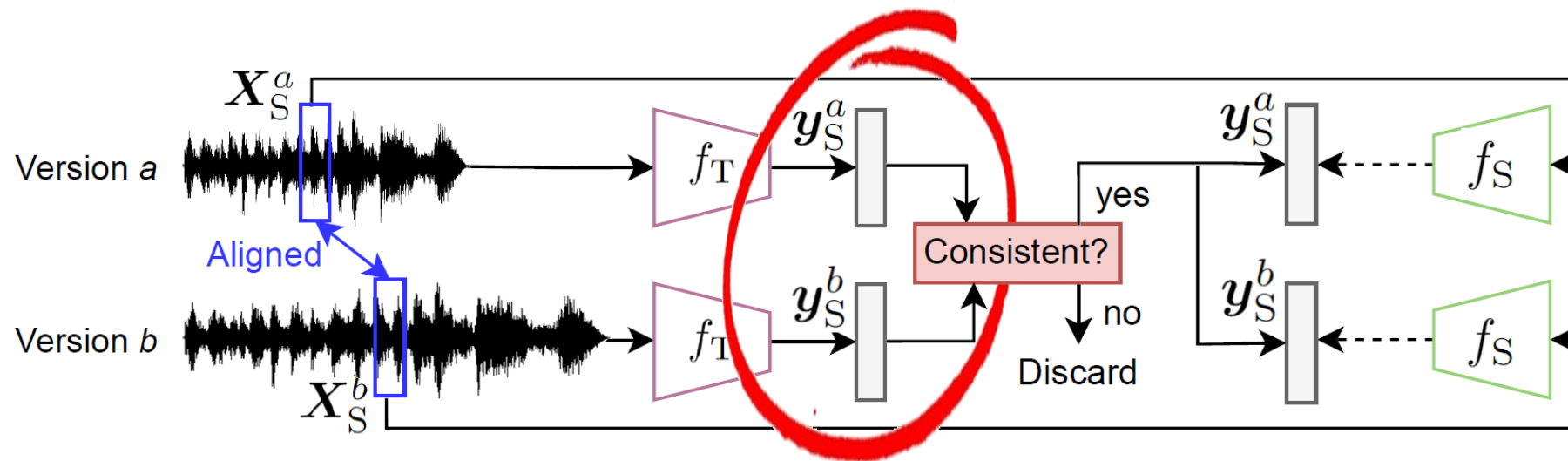
fp f p VI VI Br

Str Br Vc Kb

cresc. f

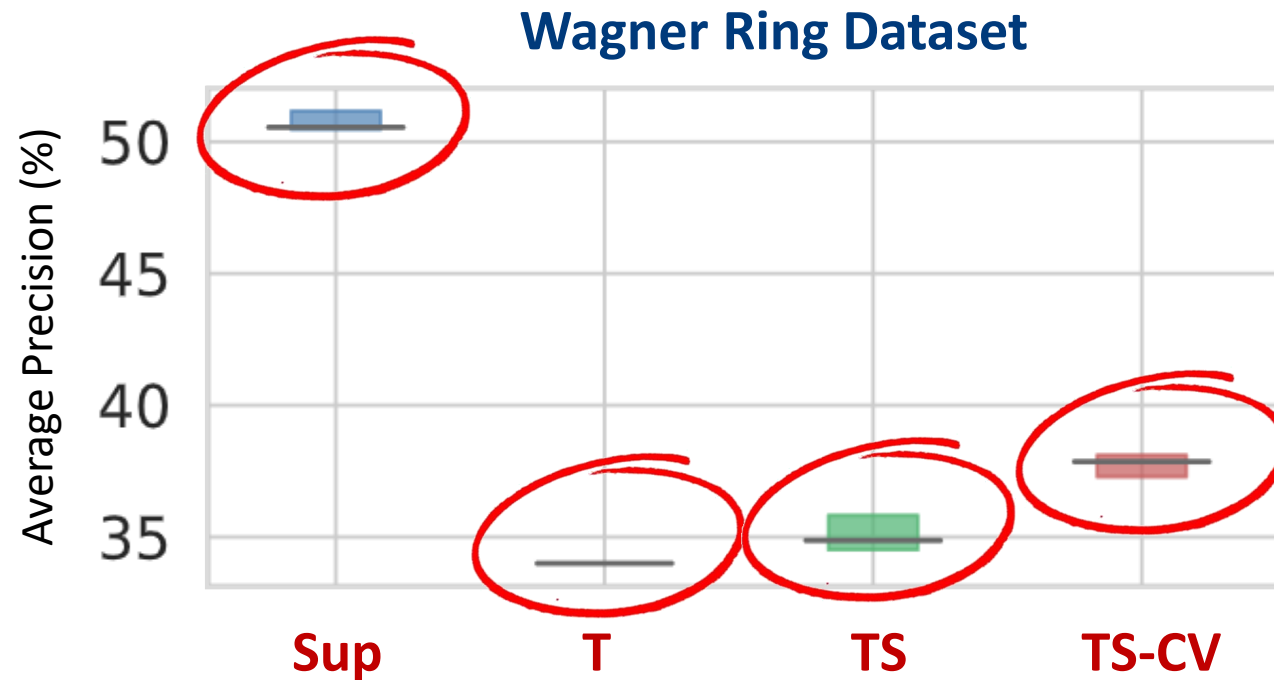
Proposed Strategy: Exploiting Cross-Version Consistency

- **Two versions** (unlabeled) of the target domain data (**performances of the same composition**)
- **Versions** (audio) should obtain the **same pitch labels** at corresponding positions
- **Align** versions and **compare pseudo-labels**
- **Enforce cross-version consistency** for teacher-student learning



Results

- Model: Conv ResNet (4.8M parameters)
- Teacher domain: **Piano solo** performances (MAESTRO dataset), AP=89.9%
- Student domain: **Expressive operas (singing and orchestra)**



Conclusion

- Enforcing **cross-version consistency** for pseudo-labels helps to **improve the domain transfer**
- Future work: **Extension to other scenarios** with multiple versions (speakers in ASR, writers in HTR, ...)

Poster Session:

Thursday 10:45 am – 12:45 pm

Paper



Source code

