

# The Challenge of Reconstructing Digits in Music Scores

---

Alexander Pacha

Enote GmbH, Berlin, Germany  
23.07.2019

[alexander.pacha@enote.com](mailto:alexander.pacha@enote.com)






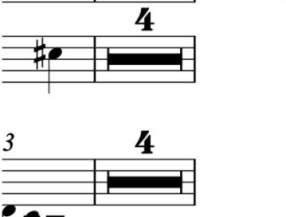
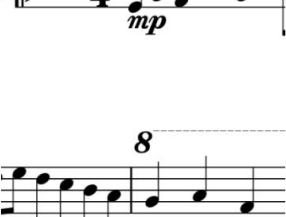
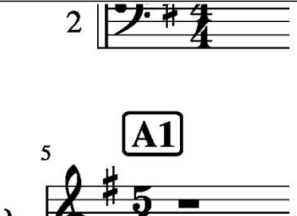
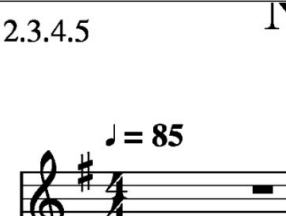



# Why Digits?

Digits play an essential role in music scores and need to be reconstructed correctly



- Necessary objective in any Optical Music Recognition (OMR) system as they convey crucial semantics
- I worked on this particular problem (and have no good solution yet)
- Nice example of why OMR is so difficult

# Digits in Music Scores

 <p>Basso Continuo</p>	 <p>Fingering</p>	 <p>Measure Number</p>	 <p>Multi-Measure Rest</p>	 <p>Octave Transposition</p>
 <p>Other</p>	 <p>Tempo Number</p>	 <p>Time Signature</p>	 <p>Tuplet</p>	 <p>Volta</p>

# Challenges

- Find all digits and correctly recognize their value
  - Primarily a computer-vision Problem
  - Deep-Learning-based approach showing promising results
- Assigning digits to the right class
  - Computer-vision and musicological problem
  - Requires knowledge about musical rules and the visual and musical context

The image displays three musical notation examples illustrating digit recognition challenges:

- Example 1:** A treble clef staff with a key signature of one sharp (F#). The first measure contains a quarter note on G4 and a quarter note on A4. A red box with the digit '3' is placed above the first measure. The label "3 / Measure Number" is written in red above the staff. The text "Vo." is written to the left of the staff.
- Example 2:** A treble clef staff with a whole note on G4. A red box with the digit '5' is placed above the note. A slur covers the next three notes: a quarter note on A4, a quarter note on B4, and a quarter note on C5. A red box with the digit '3' is placed above the third note. The label "5 / Fingering" is written in red above the first note, and "3 / Fingering" is written in red above the third note.
- Example 3:** A treble clef staff with a triplet of eighth notes on G4, A4, and B4. A red box with the digit '3' is placed below the first note. The label "3 / Triplet" is written in red below the first note. A second triplet of eighth notes on C5, B4, and A4 is shown. A red box with the digit '3' is placed below the first note of this second triplet. The label "3 / Triplet" is written in red below the first note of the second triplet.

# Challenges

**Allegro**

*p* *legato*

- Missing digits
  - Musicians are expected to understand rule of good continuation
  - Primarily a musicological problem
  - Maschine must understand underlying rules and resort to most plausible explanation
- Ambiguities
  - Digits can have multiple plausible explanations
  - Primarily a musicological problem
  - Maschine must understand underlying rules and resort to most plausible explanation

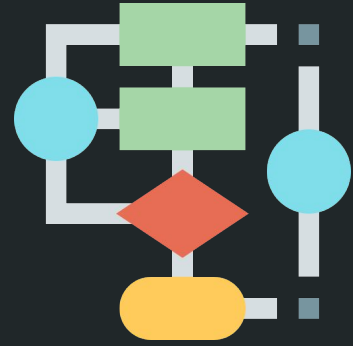
3 / Fingering & Tuplet

3 2

3

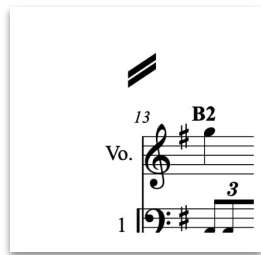
da steht ein Linden-baum;

# Digit Classifier with Deep Learning



# Digit Classification

- Digit classification can be formulated as image classification problem
  - Extracting regions of interest with digits in its center
  - Running a state-of-the-art image classification neural network (e.g., ResNet50)
- Additional information can be provided
  - The depicted digit (0-9)
  - A-priori probability of classes and/or combinations of values and classes
- Trained on dataset with over 10.000 synthetic samples and 7.000 manually annotated real samples

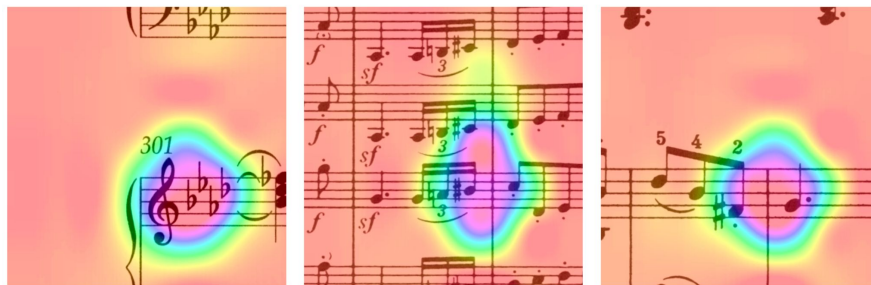


# Results

- Training performance near 100% accuracy
- Real-world performance less than 60% accuracy

Potential reasons:

- Missing attention: network classified entire image instead of digit in center
- Synthetic samples did not match real-world distribution
- Contextual (musical) information is missing
- Ambiguities force network to approximate most likely case





# Summary

- Classifying digits is a crucial step in any OMR system
- Purely image-based approaches are insufficient
  - Highly context-sensitive
  - Implicit knowledge that needs to be deduced from previous bars
  - Ambiguities need careful examination (even by humans)
- Machine-learning-based approaches might provide a good initial classification, but must fuse with contextual information
  - Make sure that your network learns the right thing

# Icon credits

Free icons from [Flaticon](#):

- [https://www.flaticon.com/free-icon/algorithm\\_1628175](https://www.flaticon.com/free-icon/algorithm_1628175)
- [https://www.flaticon.com/free-icon/asking\\_900415](https://www.flaticon.com/free-icon/asking_900415)