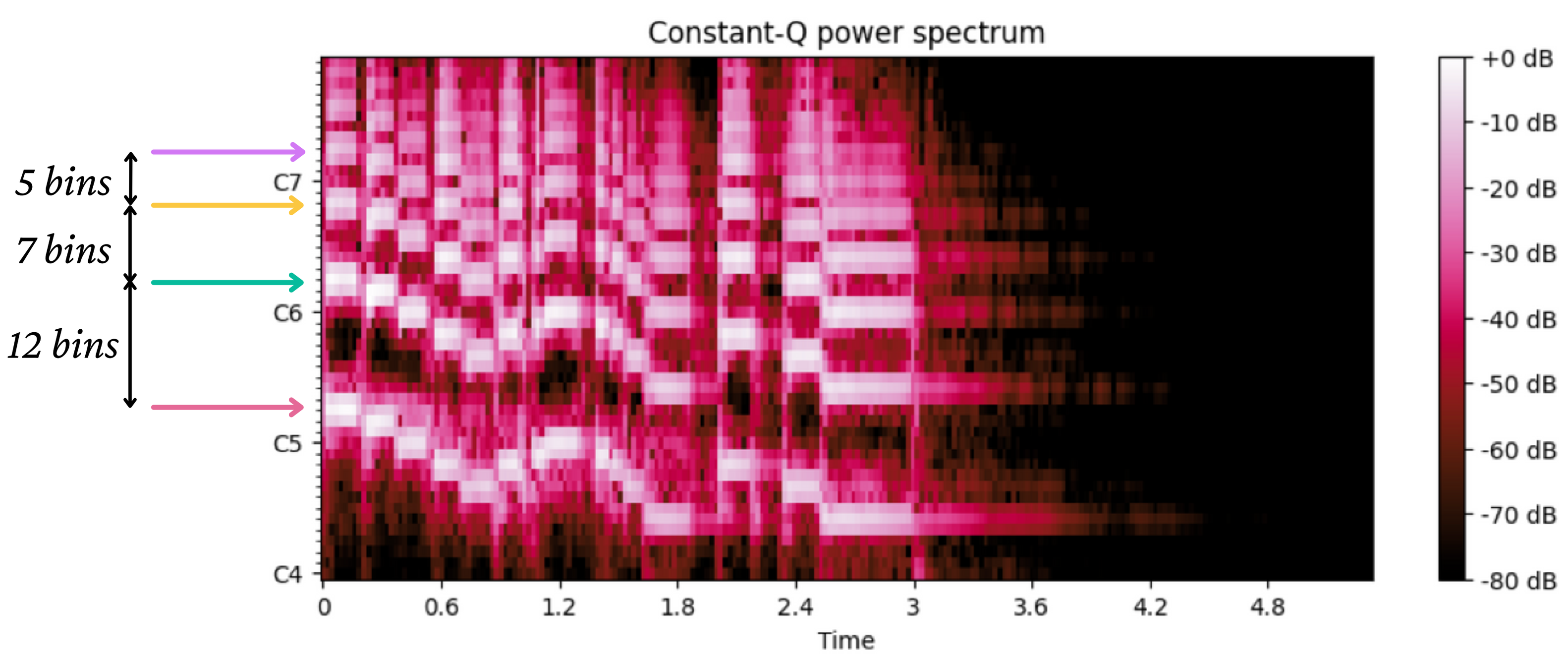
# *TriAD*: Capturing harmonics with 3D Convolutions

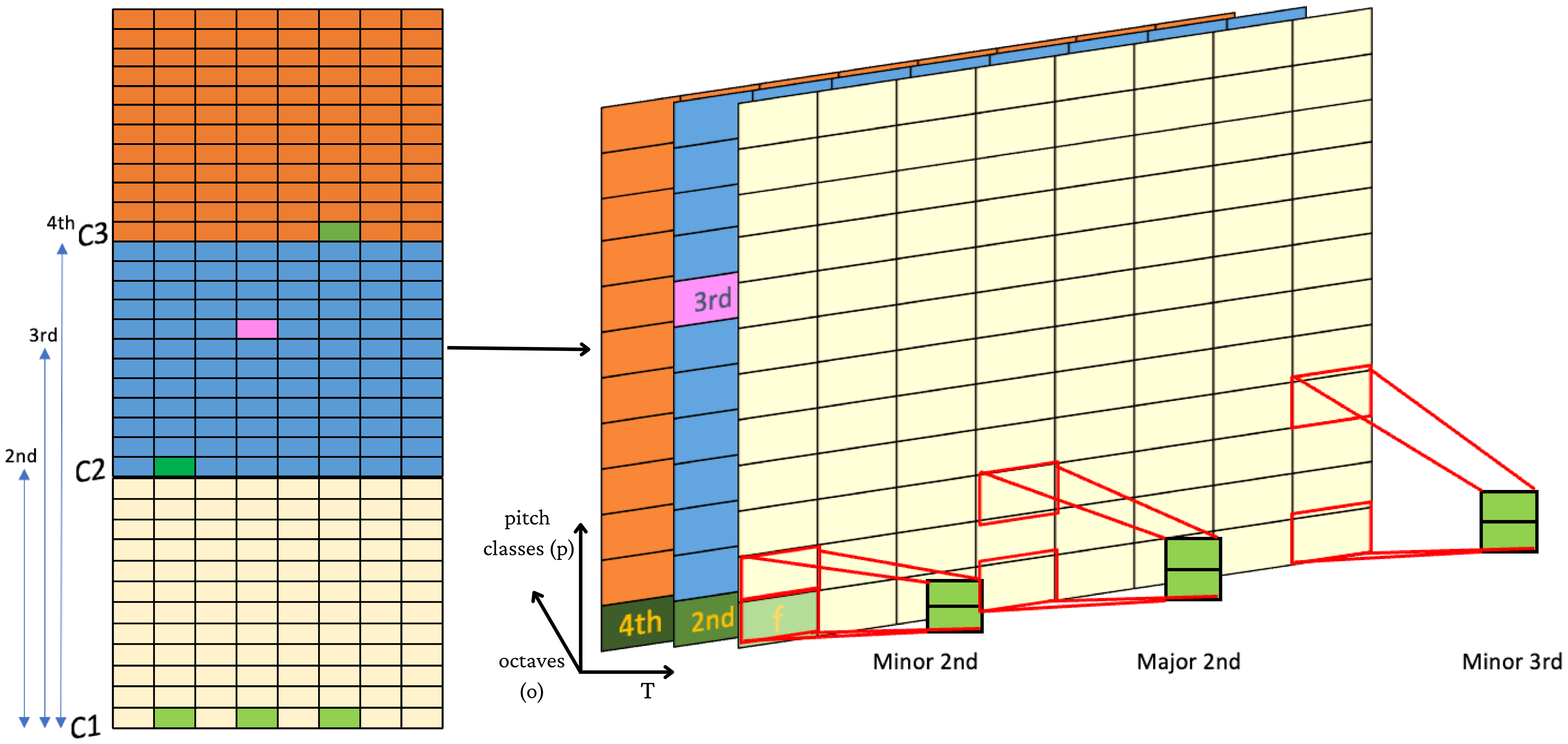
HUAWEI#
Universitat Pompeu Fabra Barcelona ♭

*Miguel Pérez*#♭
*Holger Kirchhoff*#
*Xavier Serra*♭

Deep learning automatic music transcription systems outperformed previous ones fully based on manual feature design, at the cost of being computationally expensive. New trends move towards smaller models that maintain such results by embedding musical knowledge in the network architecture. We present TriAD, a convolutional block that achieves an unequally distanced dilation over the frequency axis, allowing our method to capture multiple harmonics with a single yet small kernel, in contrast to existing methods.




- *When a note is played, the fundamental frequency sounds along its harmonics: 2nd, 3rd, 4th, etc.*
- ***Automatic transcription systems** uses that pattern to obtain accurate pitch information.*
- ***The challenge**: harmonics are not equally distanced over the frequency axis.*




- The input to the network must be a log-frequency representation of the spectrum.
- The image above displays the case of a CQT spanning $o$ octaves with $p$ pitch classes across $T$ frames.

- The CQT is split into an octave/pitch spectrogram.
- 3D are kernels dilated at the pitch-class dimension. Certain pitch-class intervals are associated with certain harmonics.
  - E.g. 2nd and 4th harmonics are octaves; 3rd and 6th harmonics are perfect fifths.
- The output of the kernels convolutions get aggregated.

- HPPNet [1] as the reference model.
- Tested multiple different harmonic blocks

- Training on MAESTRO [2]
- Evaluation on MAESTRO and MAPS [3]

- SOTA with less parameters.
- Better correlation with harmonic information (See the table below)

| | Major third | | Perfect fifth | | Minor second | | Major seventh | |
|---|---|---|---|---|---|---|---|---|
| ***Block*** | ***MAESTRO*** | ***MAPS*** | ***MAESTRO*** | ***MAPS*** | ***MAESTRO*** | ***MAPS*** | ***MAESTRO*** | ***MAPS*** |
| TriAD (Ours) | **90.14%** | **71.58%** | **90.23%** | **71.98%** | 83.16% | **68.53%** | 83.36% | **69.19%** |
| HD-Conv [1] | 84.89% | 69.96% | 85.98% | 70.50% | **84.23%** | 67.86% | **84.79%** | 68.69% |

[1] Weixing Wei et al. (2022) HPPNet: Modeling the Harmonic Structure and Pitch Invariance in Piano Transcription. Proceedings of the 23th ISMIR.
[2] Curtis Hawthorne et al. (2019) Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. Proceedings of the 7th ICLR.
[3] Valentin, Emiya et al. (2009). IEEE Transactions on Audio, Speech, and Language Processing.