AugmentedNet: A Convolutional Recurrent Neural Network for Automatic Roman Numeral Analysis with Improved Data Augmentation

Néstor Nápoles López[1], Mark Gotham[2], and Ichiro Fujinaga[1]
nestor.napoleslopez@mail.mcgill.ca mark.gotham@tu-dortmund.de ichiro.fujinaga@mcgill.ca

[1] McGill University / CIRMMT, Montreal, Canada
[2] Technische Universität Dortmund, Germany

One of the most common ways to analyze a piece of tonal music is through Roman numeral analysis. In this paper, we present the *AugmentedNet*, a convolutional recurrent neural network that improves the automatic prediction of Roman numeral analysis labels. The network is characterized by a novel representation of pitch spelling, a separation of bass and chroma inputs into independent convolutional blocks, and the layout of the convolutional layers in each block (see Figure 1). The network is enhanced by a *greater number of tonal tasks* to solve simultaneously and *synthetic training examples* for data augmentation. The *additional tonal tasks* (bottom-right side of Figure 1) strengthen the shared representation learned through multitask learning. The *synthetic training examples* consist of "new" scores, which are artificially generated from the chord annotations and texturized with simple patterns, such as an *Alberti bass* figure.

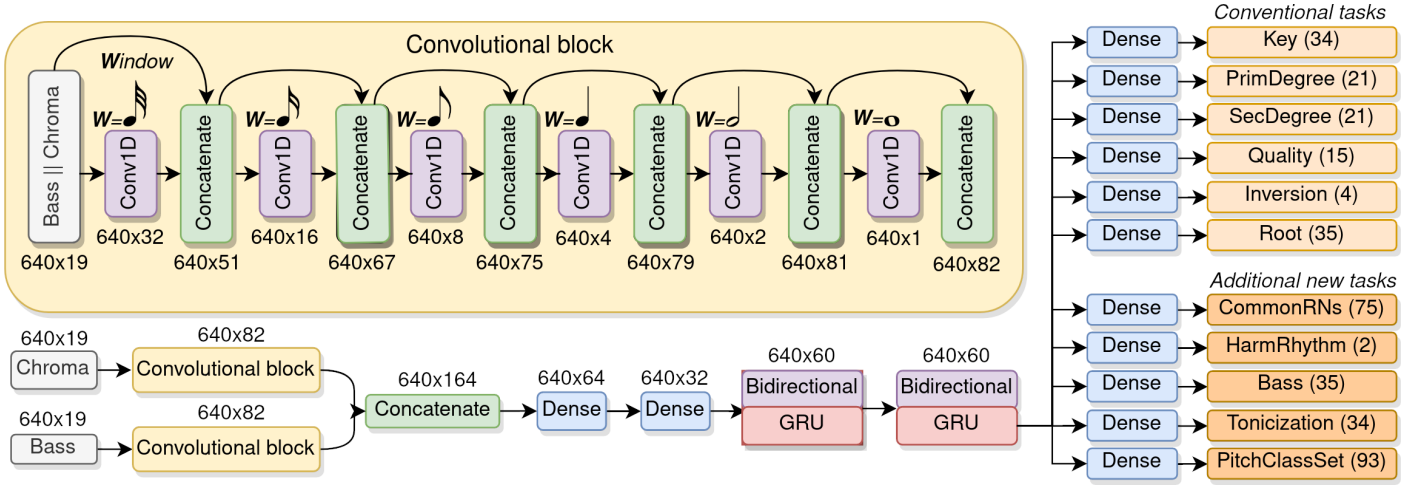

Fig. 1. *AugmentedNet.* The bass and chroma inputs are processed through independent convolutional blocks and then concatenated. Both convolutional blocks are identical and expanded on the top of the figure. A convolutional block has six 1D convolutional layers. Each layer doubles the length of the convolution window and halves the number of output filters. On the right, the MTL layout with eleven tasks. Each task indicates the number of output classes in parentheses.

We trained and evaluated the network using six datasets: Annotated Beethoven Corpus (ABC), Beethoven Piano Sonatas (BPS), Haydn "Sun" Quartets (HaydnSun), Theme and Variation Encodings with Roman Numerals (TAVERN), When-in-Rome (WiR), and the Well-Tempered Clavier (WTC). In our experiments, we found that the configuration of the network that used *synthetic training examples* and *additional tonal tasks* was the best-performing configuration in average, across the six datasets. We also found that this configuration outperformed two state-of-the-art models in automatic Roman numeral analysis [1] [2]. For more information about the experiments, please check our ISMIR paper [3]. For accessing our source code, preprocessed datasets, and experiment logs, please visit: `https://github.com/napulen/AugmentedNet`.

Acknowledgements

This research has been supported by the Social Sciences and Humanities Research Council of Canada (SSHRC), the Fonds de recherche du Québec–Société et culture (FRQSC), and Compute Canada (https://www.computecanada.ca/).

References

1. Chen, T.P., Su, L.: Attend to chords: Improving harmonic analysis of symbolic music using Transformer-based models. Transactions of the International Society for Music Information Retrieval **4**(1) (2021)
2. Micchi, G., Gotham, M., Giraud, M.: Not all roads lead to Rome: Pitch representation and model architecture for automatic harmonic analysis. Transactions of the International Society for Music Information Retrieval **3**(1) (2020)
3. Nápoles López, N., Gotham, M., Fujinaga, I.: AugmentedNet: A Roman numeral analysis network with synthetic training examples and additional tonal tasks. In: Proceedings of the 22nd International Society for Music Information Retrieval Conference. pp. 404–411 (2021)