THE 2025 KG MUSIC STRUCTUE ANALYSIS SYSTEM

Chunzhi Xiao
KG Music Apollo Lab
xiaochunzhi@kugou.net

Jingxuan Wan
KG Music Apollo Lab
wanjingxuan@kugou.net

Chuanyi Chen
KG Music Apollo Lab
chenchuanyi@kugou.net

ABSTRACT

This paper describes KGStruct, our submission to the MIREX 2025 Music Structure Analysis Task. We present a method that addresses the limitations of existing datasets and models by focusing on high-quality, multi-faceted data and a specialized model architecture. Our system utilizes a meticulously annotated, large-scale dataset that introduces new structural label types to capture the diversity of modern music, including complex electronic tracks. We adapt the All-in-One[2] as our backbone, modifying it to remove the downbeat prediction task and adding a dedicated electronic music prediction branch. This specialized branch significantly improves the recall of electronic-specific labels, which are often underrepresented in standard datasets. We demonstrate that our approach, combining a rich, high-quality dataset with a tailored architecture, leads to a substantial improvement in music structure prediction accuracy, particularly for complex and genre-diverse music.

1. INTRODUCTION

Accurately understanding the structure of music is a fundamental challenge with broad applications, from music recommendation and discovery to automated production and remixing. Traditional methods often struggle with the complexity and diversity of modern music, particularly for electronic and non-traditional genres. Existing datasets often lack the depth and quality needed to train models for these nuanced tasks, leaving a significant gap in the field.

To address these challenges, we present KGStruct, a novel approach for music structure analysis. Our system is built on a high-quality dataset and a slightly modified version of the All-in-One[2]. This combination allows for more precise and versatile structural predictions.

2. METHOD

2.1 Models

We use a slightly modified version of the All-in-One[2] model as our foundation. The original model was designed for multiple tasks, but for our purposes, we made two key changes:

Removal of Downbeat Prediction: We removed the downbeat prediction task. When our dataset was smaller, we found that including beat and downbeat prediction significantly helped prevent overfitting and improved the temporal accuracy of boundary predictions. However, as the dataset grew substantially, predicting beats still provided a positive effect, but adding downbeats did not offer any noticeable additional benefit. Furthermore, downbeat annotation is very time-consuming, so we removed this task to improve annotation efficiency without impacting the prediction quality for boundaries and segments.

Addition of Electronic Dance Music Prediction Branch: A major challenge in our dataset is the significant imbalance, with electronic dance music accounting for only 2% of the total tracks. This data scarcity makes it difficult for a single model to accurately predict specific electronic dance music-related labels, which often suffer from low recall. To overcome this, we added a dedicated branch to the model that predicts whether a track is electronic dance music. While directly predicting the three electronic dance music-specific labels is challenging, a binary classification (electronic dance music vs. non-electronic dance music) is much more reliable. When the model identifies a track as electronic dance music, this specialized branch intervenes to adjust the probabilities of the electronic dance music-related labels, making them more likely to be recalled. This improvement significantly boosted the structural prediction accuracy on our internal test set.

2.3 Dataset

A major bottleneck in music information retrieval is the scarcity of high-quality, labeled data. Most datasets focus on traditional verse-chorus structures, ignoring the complexity of other genres. Our approach tackles this head-on by building a new, diverse, and meticulously annotated dataset of approximately 6,000 songs. Our dataset is designed for broad coverage, encompassing a wide range of genres, including pop, rock, folk, various sub-genres of electronic dance music etc. The catalog spans multiple global markets, including Chinese, Japanese, Korean, European, American, and Latin American.

Beyond common structural labels like verse and chorus, our dataset includes more granular annotations to capture the nuances of diverse musical styles. For electronic music, we specifically included labels for the Break-build-drop system. Additionally, we incorporated detailed sub-types such as Pre-Verse and Post-Chorus. However, for training purposes, we mapped and merged the labels according to the rules described in Wang et al. (2022) [1].

3. REFERENCES

[1] Wang, J. C., Hung, Y. N., & Smith, J. B. (2022, May). To catch a chorus, verse, intro, or anything else: Analyzing a song with structural functions. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 416-420). IEEE.

[2] Kim, T., & Nam, J. (2023, October). All-in-one metrical and functional structure analysis with neighborhood attentions on demixed audio. In 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) (pp. 1-5). IEEE.